University of Zurich[UZH]

Evaluating transportation mode detection methods using multiple datasets and sensors

GEO 511 Master's Thesis

Author
Federica Corti
14-711-766

Supervised by
Dr, Lindsey Conrow (lindsey.conrow@geo.uzh.ch)
Hoda Allahbakhshi

Faculty representative
Prof. Dr. Robert Weibel

30.01.2020
Department of Geography, University of Zurich

University of Zurich[UZH]

Department of Geography

Evaluating transportation mode detection methods using multiple datasets and sensors

GEO 511 Master Thesis

Author:
Federica Corti
14-711-766

Supervised by
Prof. Dr. Robert Weibel
Dr. Lindsey Conrow
Hoda Allahbakhshi

Faculty representative
Prof. Dr. Robert Weibel

30th of January, 2020
Department of Geography, University of Zurich

Abstract

Mobility is fundamental for our daily lives and shapes the way we move through space. Especially with increasing age, the overcoming of distances which is important for health and social interaction is complicated. Following the increasing ageing of society in Switzerland, the MObility, Activity and Social Interaction Study (MOASIS) is executed by the University of Zurich. In this thesis, the focus is put on the mobility part, with the classification of the data into seven transportation modes, using the GPS and accelerometer sensors. The goal is to quantify the difference in classification performance based on the different sensors used and with different datasets for training and testing and improve the performance of the often misclassified public transportation modes in using features such as the heading change and stop time in a segment.

Different classifiers are trained using a benchmark dataset collected in a previous study and applied on selected data from MOASIS. It was found that the features having the highest importance for classification are simple statistical features describing the highest, lowest and range characteristics of speed and acceleration within a segment. When using both sensors combined, it depends on the training dataset, which features have the highest importance. Three classifiers are tested, the Random Forest algorithm, Support Vector Machine and Neural Network. It is found that the Random Forest algorithm performs best for transportation mode detection. For the classifier trained and tested with data from different datasets, the highest kappa value of 0.4484 is achieved in using features from the GPS sensor. When using training and testing data from the same data collection, the conclusions from the literature are confirmed. The use of accelerometer strongly improves the classification, resulting in an increase of kappa of over 0.4 from the classification with GPS. Towards improving the classification of private and public transportation, the use of the specific features proves useless. The best results are achieved in classifying the transportation modes hierarchically, using four different classifiers to split the classification based on the best distinguishable features.

For future research, the building of a more diverse benchmark dataset including different environments must be approached, which is more useful for training or testing classifiers. To increase the classification performance, the inclusion of additional GIS information about public transportation stops is advised.

Acknowledgments

I would like to express my appreciation to multiple people helping me with this thesis:

- First, and most important of all, thank you to my supervisors Prof. Dr. Robert Weibel, Dr. Lindsey Conrow and Hoda Allahbakhshi for the regular meetings throughout this year. The meetings were always helpful to get new points on view on encountered problems and on possibilities for the development of this thesis.
- To Dr. Oliver Burkhard for the *Shiny*-application enabling me to manually label part of the data for validation.
- To Vera Isler for the collection of the benchmark dataset and the aid in understanding the data.
- To Carmen Corti for proofreading this thesis.

Table of Contents

List of Figures

List of Tables

Glossary

EAR	Electronically activated recorder
GIUZ	Department of Geography, University of Zurich (*Geographisches Institut der Universität Zürich*)
GPS	Global Positioning System
MLP	Multilayer Perceptron
MOASIS	MObility, Activity and Social Interaction Study
NN	Neural Network
RF	Random Forest
SVM	Support Vector Machine

1 Introduction

1.1 Motivation

Mobility is a fundamental part of our daily lives and shapes the way we move through space. The choice of transportation mode can depend on many different factors, such as the purpose of the trip, the time disponibility and the availability of transportation modes. The coverage of distances is performed to satisfy different types of needs, such as physiological consisting of eating and sleeping, institutional including work or school and personal, such as pastime activities. Even though there is a general need for mobility, individual activity varies greatly depending on the life stage and the human environment (Vilhelmson, 1999).

An individual's mobility can be explored on the micro- and macro-scale using different sensors. Micro-scale mobility analyses physical activity regarding the physical movement of a person and includes the type and intensity of the movement, the number of steps and duration. Macro-scale mobility describes how the person moves in space on a larger scale, specifying the type of transportation modes used, the distance covered, and the speed travelled (Allahbakhshi & Weibel, 2017). In other words, micro-scale mobility consists of physical activity classification, whereas macro-scale mobility consists of transport mode classification. This thesis will focus on macro-scale mobility, which is detecting different transport modes. Mobility is described by successive speed and acceleration records at a specific time, that are combined into a trajectory. Depending on the similarity of features, trajectories can be subdivided into segments (Biljecki et al., 2013). As walking is defined as a mode of transportation, the walking segments detected by the macro-scale mobility analysis can be further researched on a micro-scale to determine the physical activity.

Due to the increasing overall age of our society, this is an interesting and relevant analysis, which allows monitoring physical and mental health, and well-being, that can be enhanced by physical activity, the real-world space use and a stimulating environment (Bereuter et al., 2016). In Switzerland, the overall age of the population has been increasing steadily. It is assumed that by 2030, 24% of the population will be over 65 years old, a number which increases to 28% by 2050. This age development will influence the mobility needs of the said age group regarding their specific mobility behaviour. Older adults are not considered a homogenous group, and there must be a distinction between younger older adults (up to 70 years old) and older older adults (over 70 years old), where generally the younger ones have a high degree of mobility and the older ones remain more stationary. Additionally, the behaviour between older citizens living in urban and rural areas is different (Marconi, 2008). In Figure 1.1, the absolute numbers of citizens over the age of 65 and their development is illustrated, as well as the increase of older adults over the age of 80 by a factor 2.75.

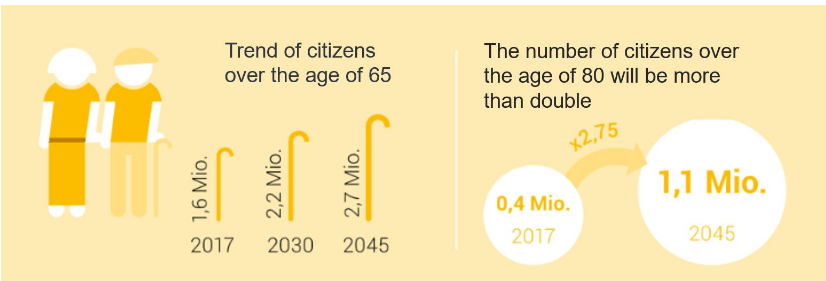

Figure 1.1: Development of older adults in Switzerland in the future (Bundesamt für Statistik, 2018).

Next to the known positive health effect of mobility, important additional benefits such as the reduction of fall risk, the improvement in mood and overall well-being and the lessening of functional decline are of increasing importance for older adults (Murphy, 2009). Besides, in order to maintain social relationships, which are essential for all age groups, but especially for older adults, mobility is required. The environment can hinder social interactions due to increasing obstacles with decreasing personal mobility (Mollenkopf, Heidrun et al., 1997). All these statements highlight the importance of studying the mobility of older adults regarding factors that enable an active and independent life.
There is an increasing number of studies showing a link between spatial mobility patterns of older adults and their level of health, well-being and independence (Fillekes et al., 2019). Most research performed on this topic has been conducted in a laboratory environment with a focus on specific health conditions. As this does not reflect the behaviour in real-life and a broader focus is desirable, it is necessary to analyse mobility indicators in real-life showing the correlation between space use, social and physical activity and cognitive well-being (Bereuter & Weibel, 2016). In addition to being tested in a laboratory setting, mobility and physical activity are often assessed by self-report measures, which are easy to acquire but subjective and influenced by different fluctuating factors, such as health, mood and cognitive ability (Murphy, 2009). Especially for older adults, questions regarding physical activity are difficult to answer, as much of the phsyical activity is light and thus unrecognised as such (Schrack et al., 2016). Therefore, in order to describe and analyse mobility, it is necessary to measure mobility-related factors objectively (Allahbakhshi & Weibel, 2018). Nowadays, the preconditions to answer questions about mobility and physical activity in the real world are far improved due to data acquisition with wearable technology, such as devices with Global Positioning System (GPS) and accelerometer sensors (Schrack et al., 2016).

1.2 Aims of this thesis

The focus of this thesis lies on macro-scale mobility, where the transportation mode classification is a fundamental part of analysing an individual's mobility. Formerly achieved in conducting travel surveys, the advance in technology facilitates and enables an objective and automated capture of mobility, using sensors such as GPS or accelerometer. On a daily basis, those sensors are carried around in devices such as a smartphone and gather objective data about the movement in space. The classifiers to analyse macro-scale mobility are built and tested using a benchmark dataset, collected within Vera Isler's Master Thesis (Isler, 2018). This is achieved in building a classification model using a benchmark dataset of labelled mobility trajectories recorded with GPS and accelerometer. The final classification is applied to a dataset collected in the real-life environment, consisting of unlabelled transportation mobility and physical activity measurements of a participant's daily life. Not only the training of the classifier is discussed, but in showing the application on an unlabelled real-life dataset, the generalizability of the classifier is tested. For the most part in literature, the implementation and training of classifiers is shown, but the further application of the classifier on an additional dataset is not discussed. In this thesis, a step in that direction is undertaken, trying to find a suitable solution to the classification problem for a specific real-life dataset.
The data, which will be classified in this thesis comes from the MObility, Activity and Social Interaction Study (MOASIS), conducted between the URPP Dynamics of Healthy Aging and the Department of Geography (GIUZ). Within the study, data is collected focusing on the analysis of mobility, physical activity, social, psychological and health aspects of older adults. Eventually, the identification of individual spatial mobility will be related to health and psychological functioning. This thesis will focus on the mobility aspect and undertake a first step in analysing the macro-scale mobility of MOASIS participants.

As the distinction between different types of motorised transport is required within this thesis, the focus lies on filling the research gap of accurate distinction between motorised transportation modes.

1.3 Research Questions

The research gaps introduced above, and summarised in more detail in Section 2.3, lead to the formulation of the following two research questions:

1. What is the classification accuracy variation for the same trajectories using only one sensor's data (GPS or accelerometer) or a combination of both sensors' data as input for the classification?
Within this thesis, it is analysed how the classifiers perform while using only position sensors (GPS), only motion sensors (accelerometer) or a combination of both. In order to analyse this, every classifier is trained three times, once using GPS-derived features, once using accelerometer-derived features and once using both accelerometer- and GPS-derived features. After the validation of these results, the difference in performance can be identified. Literature has proven that the use of both sensors provides the best classification result, but the quantification difference in classification accuracy using the different sensor combinations remains important. For the classification of data collected within MOASIS study, it will be important to quantify the trade-off between the different sensors, as the complete coverage of the trajectories by both sensors is not given. It is useful to know how much that will affect the classification accuracy. For this comparative analysis to work, it is essential to apply the classification on trajectories that are fully covered by both sensors.

2. How can misclassifications between transportation modes which have a similar signal in an urban context be reduced (car, bus, tram)?
As stated by Stenneth et al. (2011), there are similar GPS and accelerometer readings within multiple modes of transportation, such as cars, buses and trains, especially in an urban context. In order to analyse the travel behaviour reliably, the occurrence of such misclassifications must be minimised. This is tested in preventing the misclassification with a specific feature selection. This feature selection can include features that have few correlations with direct velocity for GPS readings, for example, direction change rate, velocity change rate and stop rate (Zheng et al., 2010). For the accelerometer sensor, this includes features regarding braking peaks and braking periods, where braking peaks help to distinguish along traffic and traffic moving independently. Braking periods show the intensity and volume of acceleration/braking and can thus help to distinguish between slower-moving vehicles (e.g. tram) and other motorised modalities. After the classification with these selected features, the occurrence of the problem will be visible, and the necessity of additional solutions discussed.

1.4 Structure of Thesis

After the general introduction into the topic, Chapter 2 provides the related work and background for the setting of this thesis. In that chapter, different transportation mode detection methods are introduced, starting from different classifiers, to the use of different sensors and different study designs. In Chapter 3, the two different datasets used in this thesis are introduced, discussing their different backgrounds and characteristics. The following Chapter 4 introduces and explains the different methods used for the transportation mode detection of the data, starting from the segmentation to the feature and classifier selection and finally the validation of the results. Chapter 5 presents the results of the different methods introduced, starting with the segmentation of the trajectories and followed by the importance of the different features in the selected classifiers. Following that, the actual results of the performed classifications are presented; this includes the use of different classifiers, different sensors and different combinations of training and testing data. In Chapter 6, the results are put into perspective, compared and discussed based on the research questions, including a critical evaluation of the performed analysis. Where in Chapter 7, this thesis will be concluded in highlighting the most critical findings and mentioning different ideas for future work.

2 Background

2.1 Introduction

Both in areas of traffic planning and transportation research not only information about general travel behaviour is relevant but also the focus on public health and information about individual travel behaviour has gained importance, both in the areas of traffic planning and transportation research (Ellis et al., 2014). This research implies the recording of travel information at different spatial and temporal granularities (Das & Winter, 2016). Specific areas interested in information about an individual's transportation behaviour are location-based services, transportation science and human geography. Location-based services have the goal to provide real-time information about location and movement, such as the real-time location of public transportation or alerting the user about his immediate environment.
Transportation science is interested in individual travel patterns in order to study what type of transportation means are important for the individual or a group of individuals in their daily routine. This type of information was formerly collected by travel diaries, leading to errors due to an overall underreporting of the trips and a decreasing response rate. In transportation science, trip detection can be based on heuristic rules, for example, in using a specific stationary time between two locations to detect a trip.
Human geography uses additional information about a trajectory in including relevant semantic information for that trip. An example is not only detecting when an individual is using a boat as a transportation mode but also knowing when an individual on that transportation mode is fishing.
These research areas have very different requirements on the analysis of trajectories, for location-based services; there is a need for a fast response, which comes at the cost of the correct segmentation of the trajectory. Where for transportation science, the accurate segmentation of the trajectory is more important than the time to get that information. For human geography, the semantical enrichment of segments in a trajectory is of higher importance (Prelipcean et al., 2016).

Recently the broad availability of devices used to record GPS positions has allowed researchers to objectively gain information about an individuals' movement in space (Ellis et al., 2014). Different sensors are useful to record different characteristics of transportation modes.
Current research focuses on data collected using smartphones, as they contain both onboard positioning sensors and Inertial Measurement Units (IMU). In using smartphones, near-real-time transportation mode detection is a recent development (Das & Winter, 2016). The broad availability of location-enabled devices and the general acceptance of recording personal data, helped scientists to obtain more data and solved the problem of research hindering through data insufficiency, ethics and privacy scepticism (Prelipcean et al., 2016). Even though smartphones are practical because of availability, the use of other devices can be beneficial due to limited power from competing demands on smartphones (Ellis et al., 2014).

The two most frequently used sensors for transportation mode detection are the GPS and the accelerometer or a combination of both (Prelipcean et al., 2016). The GPS records the location of the device, with varying precision depending on the number of satellites that can be reached by the GPS recorder. The three-dimensional accelerometer is a device measuring the acceleration along three axes with respect to the gravitational force. For example, if the accelerometer lays still on a surface, the acceleration downward is 1g, and the acceleration along the other two axes is zero (Shafique & Hato, 2014). Additionally,

travel data collection using these sensors does not have any requirements from the participant, as there is no work involved. The low effort for the participant can lead to longer data collection campaigns, which enable the researchers to get information about multi-day travel patterns (Wu et al., 2016). The following paragraph discusses some advantages and disadvantages of the two sensors.
The GPS sensor records the position and velocity of and individual; these recordings come from the distance of the mobile phone to each of the satellites that reach the sensor. For an accurate positioning in the two-dimensional space, the sensor must have a connection to at least three satellites. Precision increases with the numbers of visible satellites (Nikolic & Bierlaire, 2017). As soon as enough satellites are in view, the data collected using GPS benefits from a high spatial and temporal resolution (Wu et al., 2016). The most significant disadvantage in using GPS is the loss of signal, as GPS does not work indoors and the precision is reduced in urban environments, due to the satellite signals either being reflected or occluded by buildings and creating so-called urban canyons ((Nikolic & Bierlaire, 2017); (Fillekes et al., 2019)). In densely-built central business districts, the urban canyon effect can be particularly distinct, a place where much data could be recorded and accuracy would be needed for a large scale travel survey (Gong et al., 2012). The need for an unobstructed view of satellites automatically excludes some transportation modes, such as subway, from being recorded by the GPS. Even when the user is in a public or private vehicle, the signal can be lost when not close enough to a window (Hemminki et al., 2013). As a second disadvantage, the use of GPS sensors comes at a great power cost, as it is the most power-consuming localisation technique for mobile computing. Especially when using a smartphone to record travel behaviour, the battery life is reduced significantly ((Nikolic & Bierlaire, 2017); (Hemminki et al., 2013)). Additionally, when the fine-grained distinction of motorised transportation modes is required, the current GPS solutions have a modest performance (Hemminki et al., 2013).
The accelerometer measures the acceleration along three axes with respect to gravitational force. Transportation studies have found that the different movements generated by an individual while performing different activities can be well recorded by an accelerometer (Nikolic & Bierlaire, 2017). Due to the fine-grained resolution, it allows for detailed information about the movement of the sensor, which is generally applied to detect physical activity (Shafique & Hato, 2014). This detailed information about the movement enables a distinction of different motorised transportation modes (Hemminki et al., 2013). The key feature that speaks for the use of the accelerometer sensor is the low energy consumption, which allows for continuous transportation mode recordings (Nikolic & Bierlaire, 2017). Another big advantage is that the accelerometer does not need external signal sources to record information (Hemminki et al., 2013). Unlike the GPS, which may take some time to connect to enough satellites, the accelerometer does only need little starting time (Wang et al., 2010). The disadvantages of the accelerometer signal are to determine the influence of the orientation of the sensor and to distinguish that information from the movement behaviour. Relevant information can be masked by gravity and other sources of noise (Hemminki et al., 2013). One clear disadvantage is that the records can not be reused for a purpose where a spatial position is needed (Prelipcean et al., 2016).

2.2 Scientific context

Three commonly used daily mobility indicators are summarised as life space and the duration of travel with passive as well as active transportation. Life space is a concept often applied in ageing research as it summarises the extent of the older adult's day-to-day mobility. Based on this information, different health outcomes can be predicted, such as cognitive ability, physical activity and functional ability

(Bereuter et al., 2016). Different studies have reported the categorisation of transportation into different classes, where the first coarse classification lies in differentiating active and passive transportation. Active transportation modes are non-motorised transportation modes, for example, walking, running and biking. Those transportation modes are also known as soft modes, which illustrates their contribution to the reduction of congestion and pollution. Passive transportation modes are motorised transportation modes, such as car, tram, motorcycle, bus, subway, and train (Nikolic & Bierlaire, 2017).
The transportation mode detection must be performed in an automated way, as it cannot be expected by the user to tag the transportation modes in the recorded movement trajectories manually. It has different reasons, one being the lack of motivation, as the transportation mode result may not benefit the user. Another is the difficulty of the individuals to remember the exact moments where the change of mode happened, as a personal trip generally consists of multiple transportation modes. If the identification of the transportation modes is made based on only simple rules, for example, a velocity-based approach, the results are unstable, due to the difference in features from travel behaviour based on different external factors, such as weather and traffic conditions. During peak traffic, the mean velocity of a bus and a bike is challenging to distinguish. Additionally, with the velocity-based approach, it is difficult to register the different transportation modes used by one user. This shows why an approach is necessary to automatically detect the transportation modes and the transition between those (Zheng et al., 2008).
Information about the movement type of the user collecting the data can enrich context-based services provided back to the user. This knowledge is developed towards motion-profiling, which means that an action is associated with a specific motion. Examples of this concept in application with a smartphone are the automatic activation of the GPS navigator when driving a car, or the silencing of the ringtone while walking. With the development of smart home environments, the information about the motion of the individual is further aggregated with information from additional sensors from the environment (Bedogni et al., 2012). When the information of numerous users is being collected and classified, the detection of real-time traffic state is possible. Companies such as Google collect smartphone data to estimate the real-time traffic speed on the road (Stenneth et al., 2011).

The different results of transportation mode detection have relevance for many different applications, such as transportation studies, urban planning, health monitoring, computer-supported eldercare and epidemiology. Regarding the research focus, in terms of improving individual health, the focus lies on non-motorised mobility, where physical activity can be observed by the daily step-count and number of calories burned. Mostly transportation modes are classified into four categories, such as stationary, walking, bicycling and the use of motorised transportation. This has worked well for data mining operations; nevertheless, not much consideration has been given to the transfer period where there can be an overlap between stationary and other transportation modes. In most studies, after the data collection, records that have too low-velocity values are removed from the dataset, with the assumption that they can be confidently assigned to the stationary mode (Xia et al., 2014).
There have been many different approaches to automatically understand and classify human mobility, either using GPS, accelerometer or a combination of both. In the following sub-chapters, the results from different studies using different sensors are presented, as this topic is thoroughly discussed in the literature. For each sensor and sensor combination, four studies will be presented in more detail.

2.2.1 GPS

GPS sensors record the user position at a specific time. From those values speed, travel direction and more can be determined. Optimally, this information is continuous, fine-grained and describes an individual's mobility in geographical space. Many studies are conducted using GPS to classify mobility modes; some are presented in the following and summarised in Table 2.1.

Huss et al. (2014) use data collected on 12 peoples working commute, consisting of the five different transportation modes: walk, bicycle, car, bus, and train. Even though only a few participants collect the data, a large amount of observation is available with over 200'000 recorded GPS positions. With areal imagery and the participants' memory, the actual transportation modes are added after collection. In order to detect the transportation modes, only a few features derived from speed are used. The best classification results are obtained by using the 95th percentile of speed, acceleration and deceleration values on segments that are at least one minute long. Figure 2.1 illustrates the different behaviour of the transportation modes for each feature. Only the 95th percentile of speed (B) and the acceleration and deceleration (F) are determined to be adequate for distinguishing transportation modes. For the mobility pattern recognition, a non-parametric discriminant analysis is performed using the speed features mentioned above. The classification results in an accuracy of kappa = 0.73, with an increased kappa value of 0.95 when the motorised transportation modes are summarised into one category. In conducting the classification using only a few features, this approach is a valid solution for studies with a large amount of GPS data. Nevertheless, the data used for this classification is collected by a small group of people in a flat area. Using only data of the commute facilitates transportation mode detection, as it assumes that other than walking and bicycling, no physical activity is performed.

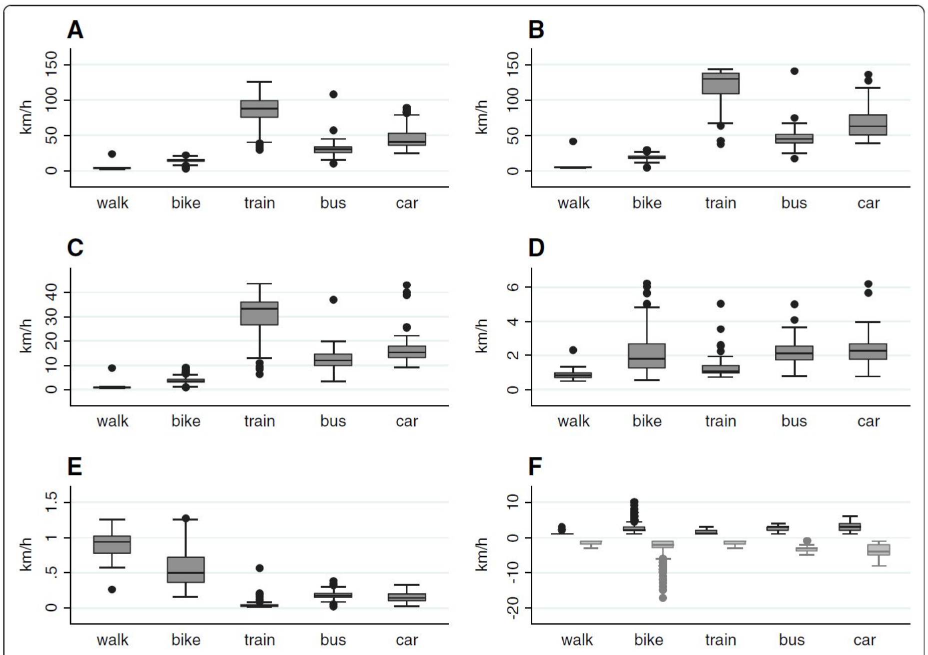

Figure 1 Speed metrics. Speed metrics per sequence for each type of transport mode, for sequences of at least one minute duration; Panels **A**: Mean; **B**: 95th percentile of speed; **C**: Standard deviation (sd); **D**: Rate-of-change (RCM); **E**: Standardized rate-of-change (RCMs) metric, **F**: Acceleration (positive values) and deceleration (negative values).

Figure 2.1: Speed metrics for the different features per transportation mode (Huss et al., 2014).

Zheng et al. (2008) propose an approach using GPS data in a supervised learning environment to identify the four transportation modes walk, drive, bus and bicycle. The used data is collected over six months by 45 users. It consists of data from 15 different cities recorded using a handheld GPS receiver and GPS phones, resulting in a total length of 20'000 kilometres. The fact that a personal trip can consist of different modes of transportation increases the difficulty of detecting the correct mode. Following

this, the first part of their work consists in detecting the change points between different modes of transportation, an approach to not only classify the transportation modes but detect the changes automatically is needed. For a trajectory containing multiple transportation modes, detection methods based only on simple rules, such as the above described velocity-based approach struggle to correctly detect transportation modes. An additional obstacle in detecting the correct transportation mode is the change of characteristic features for a transportation mode in different weather and traffic conditions. Where the traffic is heavy, the speed characteristics of a car might be very similar to those of a bicycle. Four different inference models are studied to classify the segmented trajectories, which are Decision Tree, Bayesian Network, Support Vector Machine and Conditional Random Field. With the specific segmentation method, the results prove to be worse than in using a uniform duration segmentation method, where the accuracy of 72.1% for the Decision Tree is achieved.
Gonzalez et al. (2008) use a Neural Network and assisted GPS data collected on mobile phones to detect the transportation mode automatically. Unlike humans and other algorithms, a Neural Network can detect subtle information and patterns in the data. The used data comes from 114 trips collected in Tampa, Florida. Additionally, assisted GPS data consisting of the three different modes of transportation car, walk, and bus is used. In using a Multi-Layer Perceptron, in combination with 10-fold cross-validation, the achieved overall accuracy is 91.23%. While 100% of the walking trips are predicted correctly, the accuracies for car and bus are 92.11% and 81.58% respectively. Limitations of this study are discussed by Wu, Yang and Jing (2016) and consist of the small number of trips used for training. Additionally, GPS trips are segmented manually by the user, assuring that a trajectory consists of only one mode of transportation.
Stenneth et al. (2011) use mobile phones to collect data for transportation mode detection. The data is collected by six people over three weeks and includes the following six mobility types: walk, bus, car, stationary, aboveground train and bicycle. Five different inference models, such as Bayesian Net, Decision Tree, Random Forest, Naïve Bayesian and Multilayer Perceptron, are tested to predict transportation modes. Among the tested models, the Random Forest model yields the highest precision accuracy of below 76%, with accuracies of 100% for walking and 96.8% for stationary.

Table 2.1: List of discussed transportation mode detection studies using GPS.

Name	Accuracy	Best classifier	Modes of transportation
Gonzalez et al. (2008)	91.23%	Multilayer Perceptron	Car, walk, bus
Huss et al. (2014)	Kappa: 0.73 (If all motorised transportation modes as one kappa: 0.95)	Non-parametric discriminant analysis (Statistical approach)	Bicycle, Walk and the motorised transportation modes: Car, bus, train
Stenneth et al. (2011)	Precision accuracy: <76% (93.5 % with transportation network information)	Random Forest	Car, bus, aboveground train, walking, bicycle and stationary
Zheng et al. (2008)	72.1 %	Decision tree	Bicycle, bus, car, walk

2.2.2 Accelerometer

In general, accelerometers are the most widely accepted sensors for locomotion detection (Hemminki et al., 2013). Its characteristics allow for the accurate measurement of physical activity and the recording of the number of steps (Hansen et al., 2012). Raw accelerometer data consists of records of acceleration in the three axes with the corresponding timestamp, which enables the calculation of overall acceleration to create an additional meaningful variable. In the following four studies are presented, with the most important findings summarised in Table 2.2.

Hemminki et al. (2013) present a technique for accurate and fine-grained transportation mode detection using accelerometer sensors from smartphones. The data is collected by 16 individuals from four different countries, adding up to 150 hours of data. A key characteristic presented in the paper is the possibility to identify different types of vehicles by using their acceleration and braking periods. In developing peak features, they can capture the characteristics of acceleration and deceleration patterns for different motorised transportation modes. The detection task is decomposed into a coarse-grained classification at first and subsequently followed by a fine-grained distinction. The three-stage hierarchical classification illustrated in Figure 2.2 proceeds in the following way: It starts with a kinematic motion classifier distinguishing the pedestrian transportation mode from other modalities. If no physical movement is detected, a stationary classifier is then used to determine if the individual is stationary or using a motorised transportation mode. With the motorised classifier, it is determined to which of the five motorised modalities the movement can be assigned. For the kinematic motion classifier, a discrete Hidden Markov model is applied. Using AdaBoost as an instance-based classifier, the stationary and motorised classifier conduct a segment-based classification with a simple voting scheme in enabling aggregation over the segment. With the three-stage hierarchical classifier, a mean accuracy of 82.4% is achieved in distinguishing between the seven mobility types stationary, walk, bus, train, metro, tram, and car.

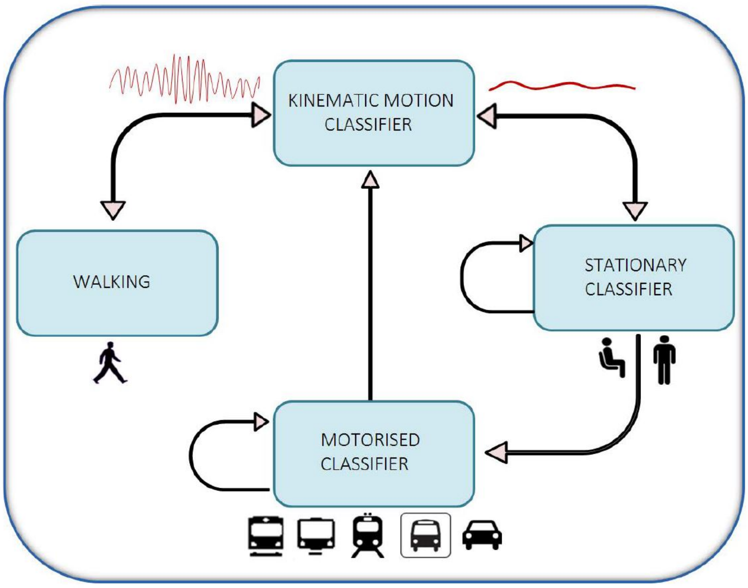

Figure 2.2: Display of the multi-step classification used for transportation mode detection, consisting of the kinematic motion classifier, the stationary classifier and the motorised classifier (Hemminki et al., 2013).

Shafique & Hato (2014) use accelerometer data to distinguish between the four transportation modes walk, bicycle, car, and train. The data is collected by 46 participants in three cities in Japan, resulting in a little over 3000 recorded trips. For the data collection, a purpose-built wearable device with GPS and accelerometer sensors is used, even though only accelerometer data is used for this study. The participant use a travel diary to collect additional information about the trip. The four different classifiers Support Vector Machine, Adaptive Boosting, Decision Tree and Random Forest are applied, and their performances compared. With an outstanding accuracy of 99.8%, Random Forest performs best out of the classifiers on this data.

Lu et al. (2018) classify the data into the five transportation modes car, bus, motorbike, walking and bike. Their work consists of two parts, first the detection of transportation mode and second the detection of different driving activities with the use of additional sensors. Focusing on transportation mode detection using smartphones, ten individuals collect about 30 hours of data. They are asked to tag the corresponding mode as one of the five transportation modes investigated in this study. Five different classification algorithms are tested to determine their impact on the accuracy results, which are Random Forest, Naïve Bayes, Decision Tree (J48), K-Nearest Neighbour and Support Vector Machine. The results show that features in the time domain are contributing most to the performance. The classification with the Random Forest algorithm returns the best result, with an overall accuracy of 97.33%, which can be explained with the selection of a random subset of features by the algorithm, thus reducing sensitivity towards correlated features.
Fang et al. (2017) study a mechanism to determine the modes of transportation using a database containing more than 8000 hours of accelerometer, magnetometer and gyroscope measurements that is collected by 224 volunteers. The data consists of ten transportation modes: still, walk, run, bicycle, motorcycle, car, bus, metro, train and high-speed rail. For the classification, the passive transportation modes are all summarised in the category vehicle and considered as a single mode. Five commonly used classification algorithms are tested: AdaBoost, Decision Tree, K-Nearest Neighbour, Support Vector Machine and Deep Neural Network. The Deep Neural Networks achieves an overall classification accuracy of 95.7 % and outperforms the other well-known machine learning methods.

Table 2.2: List of discussed transportation mode detection studies using accelerometer.

Author	**Accuracy**	**Best classifier**	**Modes of transportation**
Fang et al. (2017) (Including gyroscope and magnetometer)	95.7%	Deep Neural Network	Still, walk, run, bicycle, vehicle
Hemminki et al. (2013)	82.4%	Multi-stage hierarchal classifier (Hidden Markov Model)	Stationary, Walk, Bus, Train, Metro, Tram, Car
Lu et al. (2018)	97.33%	Random forest	Car, Bus, Motorbike, Walk, Bicycle
Shafique & Hato (2014)	99.8%	Random forest	Walk, Bicycle, Car, Train

2.2.3 Combination of GPS and Accelerometer

Transportation mode detection approaches that combine the introduced sensors are first introduced in 2004, without being evaluated until later studies considered the method an interesting approach for future transportation mode detection (Reddy et al., 2010). In the mobility analysis, two different types of sensors are favourable: motion sensors and position sensors. Transportation mode classification uses motion sensors to record vibration or oscillation characteristics, and position sensors to derive speed information (Xia et al., 2014). Generally, the use of accelerometer data is expected to improve transportation mode detection (Bedogni et al., 2012). Due to the inclusion of more fine-grained information that the one from the GPS, it enables a distinction on a smaller scale (e.g. enhanced distinction of different motorised transportation modes) (Hemminki et al., 2013). Accuracy generally increases by 10% to 20% compared to classifications that use only one sensor as input (Xia et al., 2014). A limitation to the one sensor-based records, especially for location indicators, is the non-usability of movement data due to signal outage (Fillekes et al., 2019). Here lies the strength in using multiple sensors, which is the possibility to overcome the loss of one signal (e.g. GPS in urban canyons) by using data from the other signal ((Ellis et al., 2014); (Nguyen et al., 2013); (Widhalm et al., 2012)). In the following, four studies are presented and summarised in Table 2.3, where both sensors are used to detect different transportation modes.

Reddy et al. (2010) build a fine-grained transportation mode classification system using a mobile phone, which includes both sensors. The data is classified into the five transportation modes stationary, walk, run, bike and motorised transportation. It is collected by 16 individuals, who are asked to collect 15 minutes of data for each of the five transportation modes. Multiple classification approaches are tested, such as Decision Tree, k-Means Clustering, Naïve Bayes, K-Nearest Neighbour and Support Vector Machines. The best overall result is achieved in combining the best instance classifier Decision Tree and follow it by a Hidden Markov Model, achieving a classification accuracy of 93.6%. Figure 2.3 shows the entire process that is performed in this study to classify transportation modes. An additional finding of the study is that the classification system produces accurate predictions, irrelevant of the position or orientation of the sensors while collecting the data.

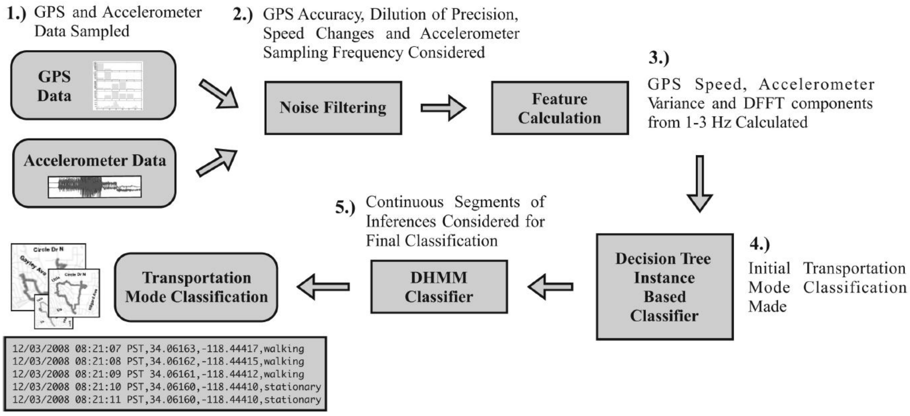

Figure 2.3: Visualisation of entire classification process (Reddy et al., 2010).

Martin et al. (2017) aim to not only achieve a high classification performance but also to reduce the data dimensionality, in order to decrease the computational cost for the classification. The developed classification algorithm can recognise five modes of transportation, such as walk, bicycle, car, bus and rail. Using smartphones, six students collect a total of 96.59 hours of GPS and 98.62 hours of accelerometer data, including the annotation of the real transportation modes. In using the K-Nearest Neighbour algorithm and the Random Forest, two of the most popular classification algorithms are tested. K-Nearest Neighbour recognises the similarity in feature characteristics for different episodes of the same transportation mode, whereas the Random Forest algorithm detects the key features that best differentiate between the modes. The Random Forest algorithm performs best with a total accuracy of 96.8%.

Ellis et al. (2014) discuss travel mode detection of five different activities, such as bicycling, riding in a vehicle, walking, sitting and standing. The focus is more on physical activity types and less on transportation mode detection, as all motorised modes are summarised as one. Two people, wearing extensive custom build-devices collected 150 hours of GPS and accelerometer data. During the experiment, it was tried to simulate real-life by collecting data in difficult conditions, such as in indoor and urban environment. Some frequently used supervised machine learning methods, such as a K-Nearest Neighbour, Support Vector Machines, Naïve Bayes, Decision Tree and Random Forest are tested, with the Random Forest algorithm providing the highest overall accuracy of 89.8%.

Xia et al. (2014) use GPS and accelerometer data collected by smart devices to classify different outdoor transportation modes, such as stationary, walking, bicycling and motorised transport. As a new approach, they classify the stationary mode further into stay and wait. Stay describes the remaining at a fixed location over a prolonged time and wait describes short stationary periods, such as waiting at a traffic light. Eighteen individuals using smartphones collect over 16'000 transportation segments. The subjects are asked to set the transportation mode flags necessary for the verification of the classifiers. The Support Vector Machine is selected as the classifier, as it is well suited for small samples, nonlinearity, regression and classification of high-dimensional patterns, resulting in a performance of 96.31%.

Table 2.3: List of discussed transportation mode detection studies using GPS & accelerometer.

Author	Overall accuracy	Best classifier	Transportation modes
Ellis et al. (2014)	89.8 %	Random Forest algorithm	Bicycling, riding a vehicle, walking, sitting and standing
Martin et al. (2017)	96.8%	Random Forest	Walk, Bicycle, Car, Bus, Rail
Reddy 2010	93.6 %	Decision Tree followed by a Hidden Markov Model	Stationary, Walking, Running, Biking, Motorised travel
Xia et al. (2014)	96.31%	Support Vector Machines	Walk, Bike, Motor, Stay

2.3 Research Gap

In general, future research will need to investigate the influence of the selected model, depending on the sensor used, such as challenges about the combination of data of different quality and sampling frequency (Das and Winter, 2016). Following the presented literature in this chapter, most studies focus on the best possible classification results of transportation mode detection using one or the combination of both sensors, such as Reddy et al. (2010) showing the good results of using both sensors. But not many studies have been performed where the goal is to detect the difference performance between the sensors, especially not in context with the use of different datasets for training and testing. The analysis performed in this thesis will quantify the difference in performance based on sensor use.
During the literature research, it becomes clear that the misclassification remains high for transportation modes with similar readings. Cars and even bicycles can have similar feature values depending on the situation and must be differentiated with high confidence from buses (Mäenpää et al., 2017). Many of the presented studies that achieve very high accuracy values summarise all passive transportation modes into one category. The results show the good distinguishability of active and passive transportation modes, but also difficulties remain for the classification into the specific passive transportation modes. One way to solve the problem of classification insecurity is to include additional data, such as transportation network information. The additional information gained using this data, is more specific information about the transportation mode, such as the particular bus/bus line that is used (Stenneth et al., 2011). Ideas to improve GIS algorithms, such as the processing of multiple-day GPS recordings and recognising repetitive patterns of travel modes or including information from the internet are proposed (Gong et al., 2012). As these approaches ask for the inclusion of more data, the goal of this thesis is to improve the classification of similar transportation modes in selecting specific features within the recorded data, without including information from third parties.
A second aspect that is striking after the literature research is a general overestimation of the accuracy due to the testing of the algorithms with data collected outside of a laboratory but under ideal conditions. When real-life mobility is tracked, there is data loss from urban canyons or indoor movement, and

there are cold start journeys where the GPS device first has to build up the connection to the satellites to determine its position. Besides, some fast mode changes can occur, or transportation modes can be collected that were not thought of in advance. These statements call for the testing of travel modes under different real-world conditions. Most of the studied literature tests the classifiers with leave-one-day out or a testing dataset cross-validation, which has been collected together with the data used for the analysis (Ellis et al., 2014). The use of data collected in the same environment leads to higher accuracy results for the tested classifier. Within this thesis, the goal is to test the translation of training classifiers using a specific dataset and to apply them on a different dataset. Eventually, this should enable the application of a pre-trained classifier on any unlabelled dataset to gain information about the modes of transportation.
In summary, these research gaps will be discussed within the scope of this thesis:

- The quantification of difference in performance depending on the sensors
- The distinction of transportation modes that are difficult to classify due to their similarity
- The translation of a classifier between training and testing data collected separately and by a different age category

3 Data

3.1 Introduction

This thesis uses two different datasets containing mobility data collected by GPS and accelerometer. In a previous Master Thesis, a benchmark dataset was built consisting of seven different modes of transportation, including bus, bicycle, car, commuter train, train, tram and walk (Isler, 2018). The goal is to train classifiers using the benchmark dataset and then apply them to obtain information about the macro-scale mobility of the unlabelled MOASIS dataset. Thus, the assumption is that the same modes of transportation are found in both datasets.
Both datasets are recorded using the same custom-built mobile device *uTrail*. It contains three different sensors, measuring the different aspects of interest for the MOASIS-study. The recorded spatial mobility is obtained by the GPS sensor, the physical activity by the IMU containing a 3-axis accelerometer and the social interaction by a microphone using the electronically activated recorder (EAR) sensor (Bereuter & Weibel, 2016). For the remainder of this thesis, only the data recorded by the first two sensors highlighted in the red frame in Figure 3.1 is used.

	Sensor	Variable
Spatial mobility	GPS	timestamp, latitude, longitude
Physical activity	IMU	timestamp, acceleration (x,y,z), step counter
Social interaction	EAR	timestamp, sound sample (30 secs)

Figure 3.1: Design and characteristics of the custom-built recording device u-Trail (Bereuter & Weibel, 2016). The sensors used in this thesis are marked by the red box.

3.2 Benchmark dataset

The benchmark dataset has been collected because of the lack of comparability in the widely discussed topic of transportation mode detection. Each researcher works with data collected individually, with different sensors and in different settings. In literature, the process of data collection is briefly discussed, usually without giving further information about the data. Typically, a subset of the data (training data) is used to train a classifier, and the rest of the data (testing data) is used to test the performance of the classifier. This approach shows the problem of transferability of classifiers; often high accuracy values are only achieved where data from the same environment is used for training and testing. In order to get unbiased results for the trained classifiers, the testing dataset should be independent. As there is no standardised evaluation procedure, the collection of a reference dataset has been undertaken, for different researchers to test their classifiers with the same dataset and obtain comparability in results (Isler, 2018).
The benchmark dataset has been collected both in Austria and in Switzerland, but only data collected in Switzerland is used for this thesis, due to the testing data being collected mostly in the same country. The collection campaign of the reference dataset is conducted using the above describe custom-built *uTrail* device. An additional smartphone with the same sensors and a smartphone used to label the trajectories precisely have also been used for the data collection. The smartphone used to annotate the trips is

carried in the participant's hand; the *uTrail* and second smartphone are either carried in a pants pocket or a bag/backpack. For the collection of the data, the user follows a predefined protocol, which defines the modes and modalities as well as the position of the devices. The transportation modes to be recorded are chosen based on the requirements for modes of transportation in the field. Different trajectories are designed to combine different public, individual motorised and active modes of transportation.
All available public modes of transportation describing daily mobility in the city of Zurich are taken into account. Motorised individual transport is covered by car data, public transport by bus, commuter train, train and tram, and active mobility by walking and cycling. Two different types of scripts for the data collection are defined, describing only one mode of transportation (unimodal) or including two means of transportation, separated by walking segments (bimodal). Unimodal trips have a duration between 10 and 15 minutes. For the recording of bimodal trips, the segments of each mode must be at least one minute long. The participants receive different instructions for the different transportation modes. When recording bicycle segments, some of the instructions are to drive on a gravel road, drive in an urban environment and to drive up/downhill. When using one of the public transportation modes, the participants are requested to sit, stand, change the seat and move around in the mode of transportation. When using a car, it is asked to record driving in different traffic situations, as well as at different speeds. The aim of these instructions is to gather real-life data from different environments and by different people but under controlled conditions. The data is collected by 12 participants from GIUZ, as well-instructed participants are required to execute the scripts. In general, the data is collected in an urban environment, focusing on day-to-day mobility. The goal of the data collection is a well-distributed, balanced set of trajectories in the urban environment of Zurich (Isler, 2018).
In the collected benchmark dataset, there is a total of 133 labelled trajectories, consisting of 44 unimodal- and 89 bimodal trajectories. In total there are approximately 25 hours of data, with a single trajectory having a length between 5 and 55 minutes. The sampling frequency is constant for the entire data collection and consists of 0.5 Hz for the GPS measurement and 25 Hz for the accelerometer. These frequencies result in different numbers of recordings per sensor, which are illustrated in Figure 3.2. It can be seen that the most often recorded transportation mode is walking, followed by car. The other transportation modes all have a similar amount of data, with commuter train having fewer records from the GPS sensor than for the accelerometer sensor, probably due to signal loss.

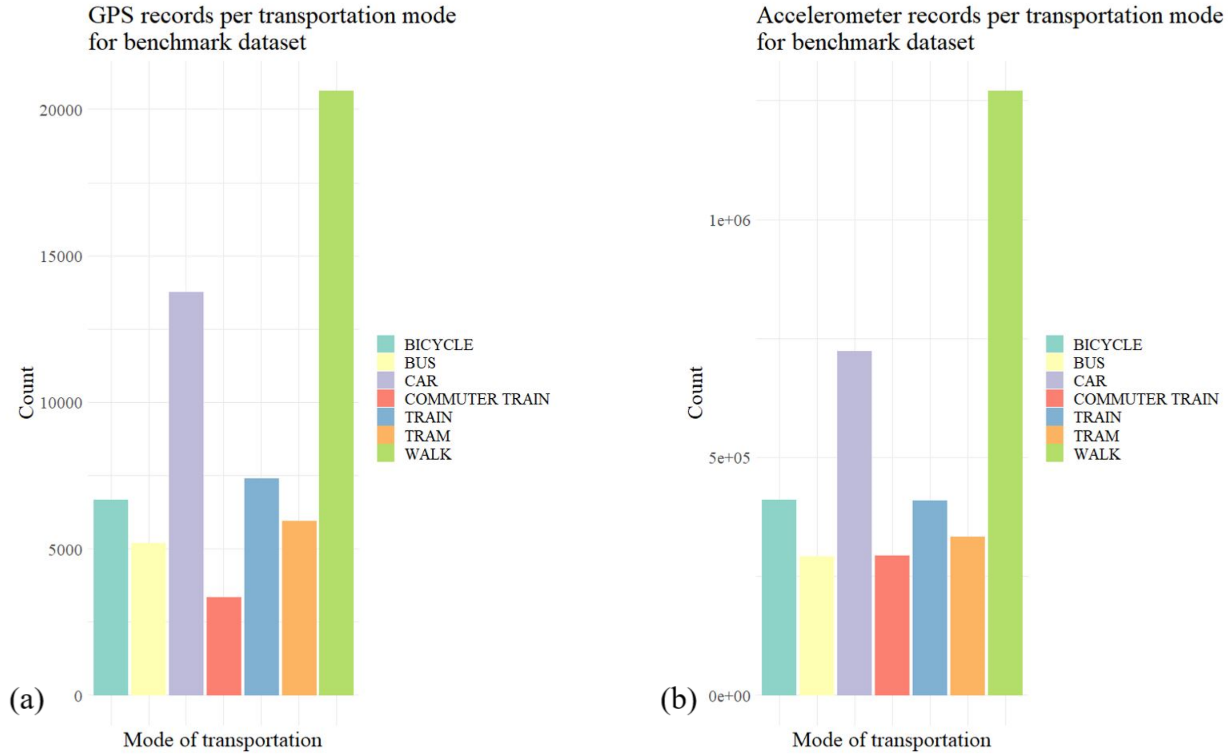

Figure 3.2: Number of records per transportation mode in the benchmark dataset, on the left side (a) for the GPS sensor and on the right (b) for the accelerometer.

In Figures 3.3 & 3.4, the environment in which the benchmark dataset has been collected is displayed on a map. It can be seen that the collection of trajectories focuses on the urban environment of the city of Zurich. There are isolated trajectories collected in the more rural area, mostly belonging to the car mode. Within the entire area of the city of Zurich, all the seven different modes of transportation were recorded.

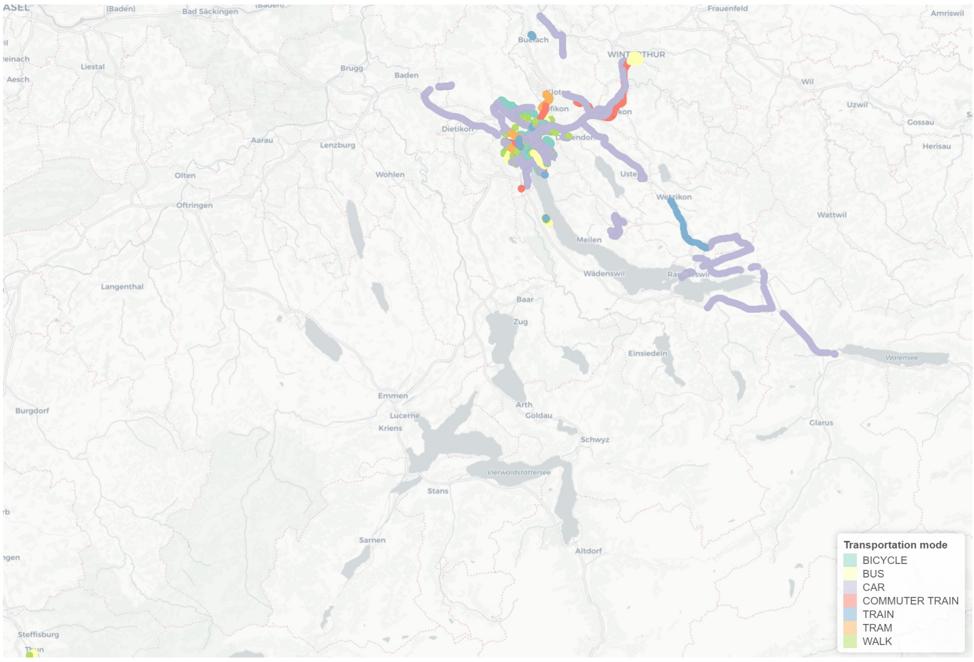

Figure 3.3: Total geographical extent of all the trajectories collected within the benchmark dataset.

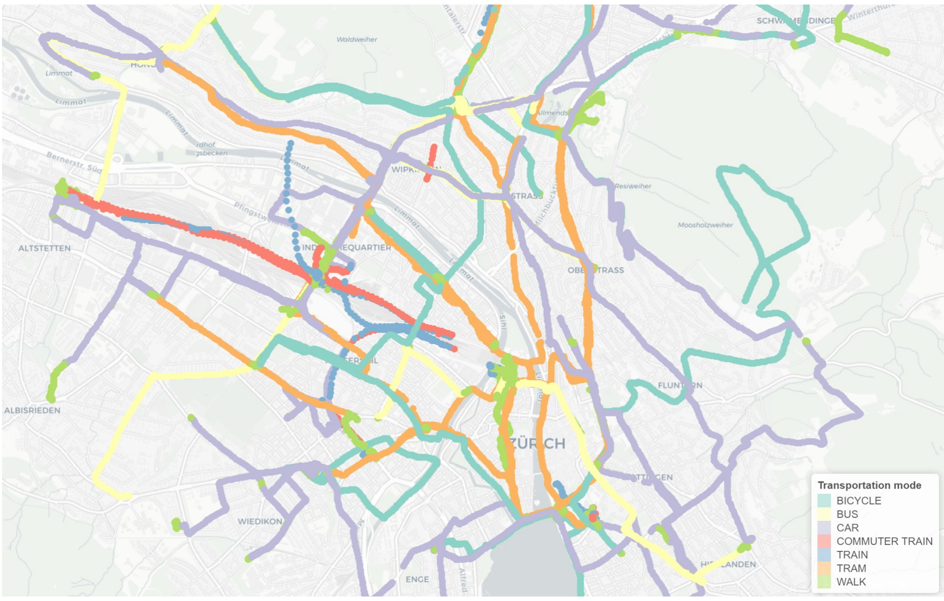

Figure 3.4: Close-up of all trajectories from the benchmark dataset collected in the urban environment of Zurich.

3.3 MOASIS dataset

The mobility data used to test the classifiers in this thesis is collected within the MOASIS study. For the entire study, a total of 164 participants over the age of 65 volunteered to carry a *u-Trail* device to

collect data regarding their mobility, physical activity and social interaction for a period of two times two weeks. Participants from the German-speaking part of Switzerland are selected for this study, with the requirements being the ability to walk without help and to have no cognitive impairments (Fillekes et al., 2019).
In the study description, the participants are asked to carry the device laterally on the hips, preferably on the belt or in a pocket. The entire data collection is split into two collection periods of each 16 days. Of the 164 participants, a total of 17 participants have no recorded valid hours, due to either not using the device or a malfunction of the *u-Trail*. Valid hours describe the time where the recording device has a valid GPS connection. As one of the goals of this thesis is the distinction of classification results based on the different sensors, it implies that for the analysis, only data with full coverage by both sensors is used. The data used for this thesis has been chosen based on the total of valid hours for the GPS per participant. For this, the valid data of both data collection periods is added up and ranked, and the 13 participants with the highest valid hours have been selected. Initially, the idea was to use the data per participant for both collection periods, but looking at the data each participant has repetitive daily mobility patterns, for the specific research performed within this thesis, it was decided to use the data of the two-week collection period with the highest valid hours per participant. A total of 14 participants at 16 days each has been chosen, resulting in the selection of 224 days and a total of approximately 2'202 valid hours. In Figure 3.5, the distribution of total valid hours for the entire MOASIS dataset is shown. Data from participants marked with the red lines has been selected for this thesis.

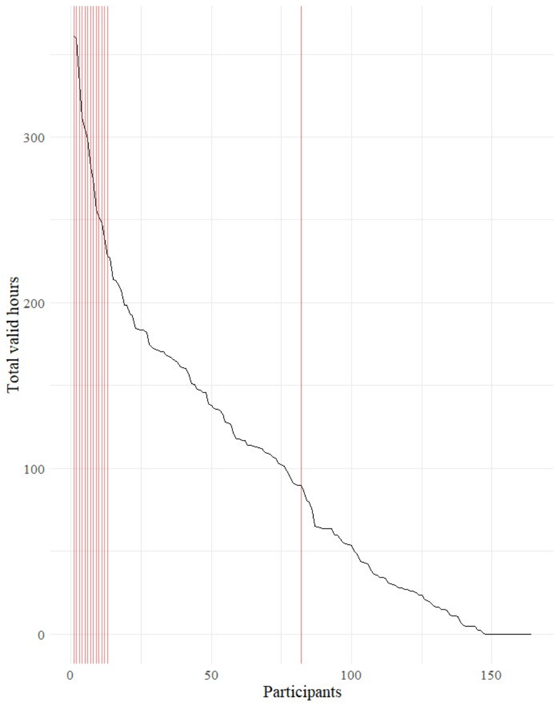

Figure 3.5: Total valid hours per participant recorded within the MOASIS study. Red lines mark the data of the participants used. The data of the 13 participants with the highest hours is selected, as well as the participant at the top of the list, which can be seen in the middle.

The sampling interval used in the MOASIS study was defined to be 1Hz for GPS and 3Hz for the accelerometer. In addition to the data collection, the participants were also asked to mark the daily self-reported indicators of life space and the time spent in each mode of transportation (Fillekes et al., 2019).

The data from the MOASIS study is not labelled; in order to assess the accuracy of the classification methods, the selected MOASIS data has been manually labelled using a *Shiny*-application (Burkhard, 2017). Figure 3.6 illustrates the labelling interface, where the trajectory is displayed on a map, with visual information about the speed and the accelerometer signal at the bottom. On the left side, the different transportation modes can be selected, and as soon as part of the trajectory is labelled the colour of the trajectory changes on the map, as well as on the speed and acceleration profiles and in the transportation mode overview.

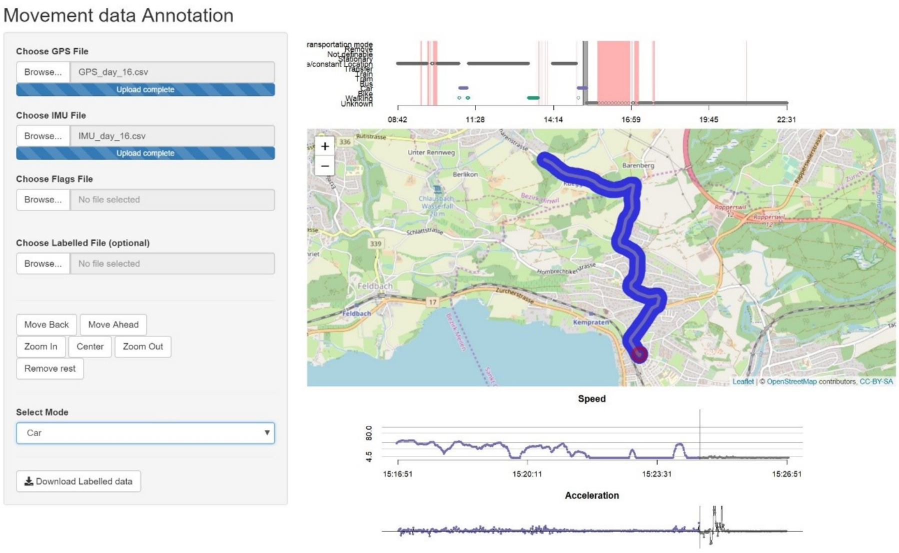

Figure 3.6: Screenshot of labelling interface used to assign the transportation modes for validation in the unlabelled MOASIS data.

Even though data with the highest valid hours is selected for the analysis, the quality of the data proves to be worse than expected. Aggregating the data in time intervals of one minute shows that the resolution of the data for both sensors was for some part of the data of around 2Hz and for a large part of the data only of one record per minute (0.02Hz). In line with the resolution of the benchmark dataset, it is determined that the MOASIS data to be used must have the resolution of at least 0.5Hz. Figure 3.7 shows the decrease in usable labelled MOASIS data due to the poor resolution. The entire amount of labelled data is illustrated in dark green, due to the low frequency only the recorded minutes in light green are used for the analysis.

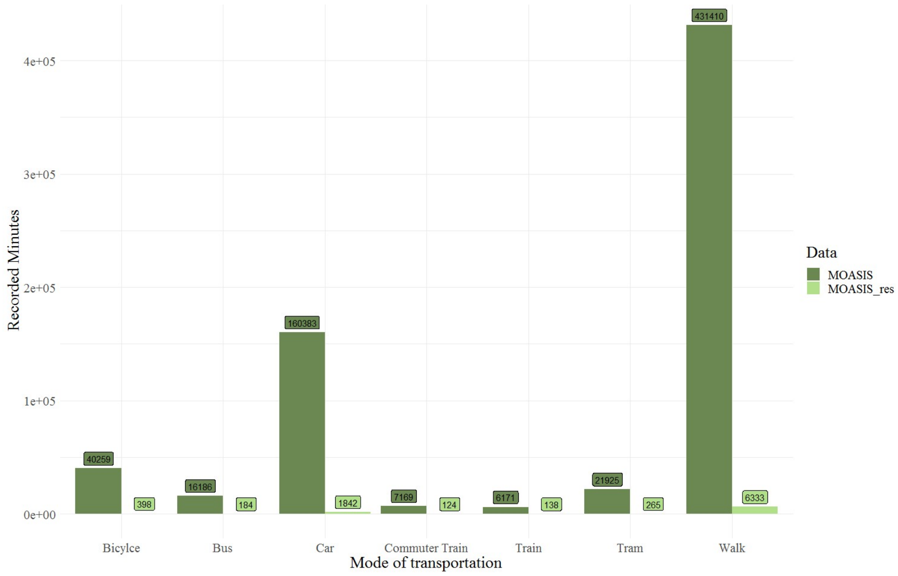

Figure 3.7: Plot showing the data labelled by minute. Dark green bars show all the labelled data per mode and light green bars the data that is used due to a resolution of at least 0.5Hz.

For the MOASIS data, only data with a full coverage by both sensors is chosen. Due to the considerable amount of data and the labelling process, it follows that the same data frequency is selected for both sensors, possibly reducing the accuracy of the accelerometer. The total amount and distribution of transportation modes for the MOASIS data can be seen in Figure 3.8. The data distribution is similar to the benchmark data; walking has the highest number of records, followed by car. The other modes of transportation have few entries, especially commuter train and train have little data. This amount of data comes from the signal loss typical with those types of public transportation. As for the number of data recorded, the MOASIS dataset has more entries than the GPS benchmark dataset.

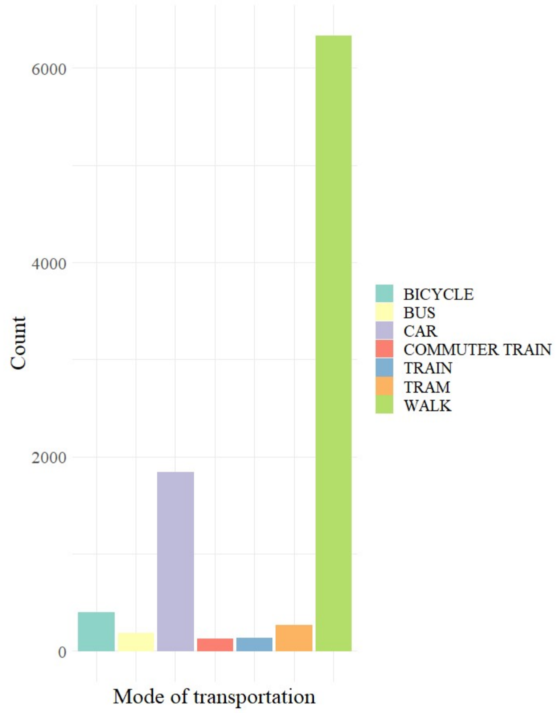

Figure 3.8: Number of records per transportation modes by minute for the MOASIS dataset.

Looking at the spatial distribution of the MOASIS data used in this thesis, the geographic extent is much bigger than in the benchmark dataset (fig 3.9). With the benchmark dataset mostly collected in the urban environment of Zurich, only a limited amount of selected data from the MOASIS dataset is collected there. Depending on the daily activities and the location of residence of the individual, the area of movement varies greatly. This geographical overview shows that the benchmark dataset is artificial in the way that it defines mobility in covering different transportation modes in a restricted environment, whereas the MOASIS data records real-life mobility of individuals that plan their activities based on subjective factors. In the overview in Figure 3.9, many longer trips are made using cars, followed by some trips with different public transportation modes.

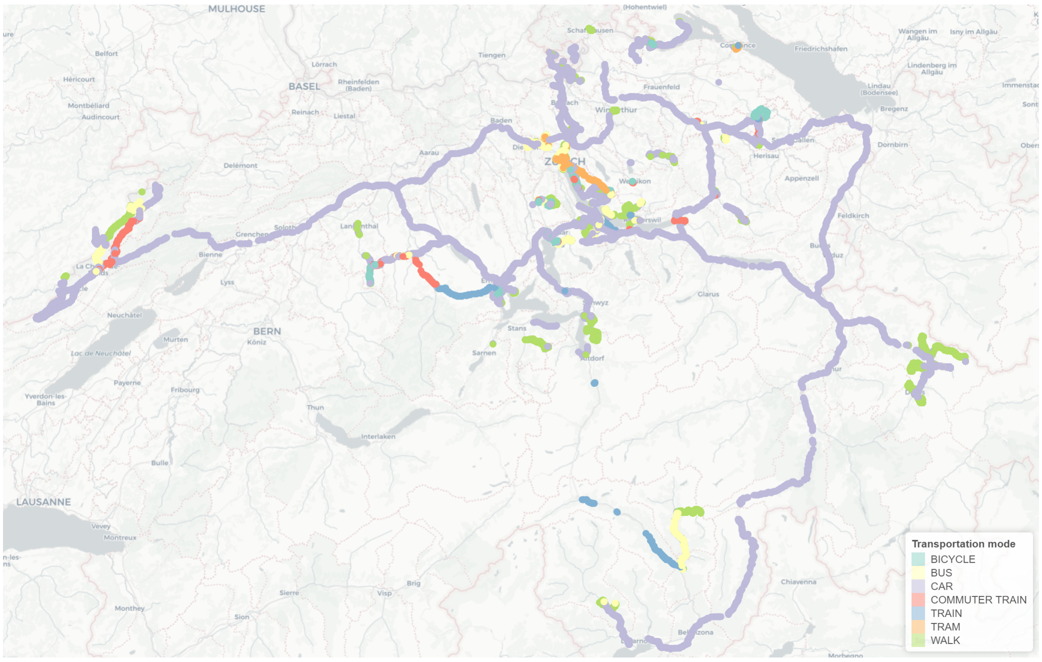

Figure 3.9: Total geographical extent of the trajectories used from the MOASIS study.

To compare the benchmark data regarding the geographic extent, a map of the greater Zurich area is shown in Figure 3.10. Not many trips are recorded in this environment; although there is a small concentration of different public transportation modes and car segments that can be detected.

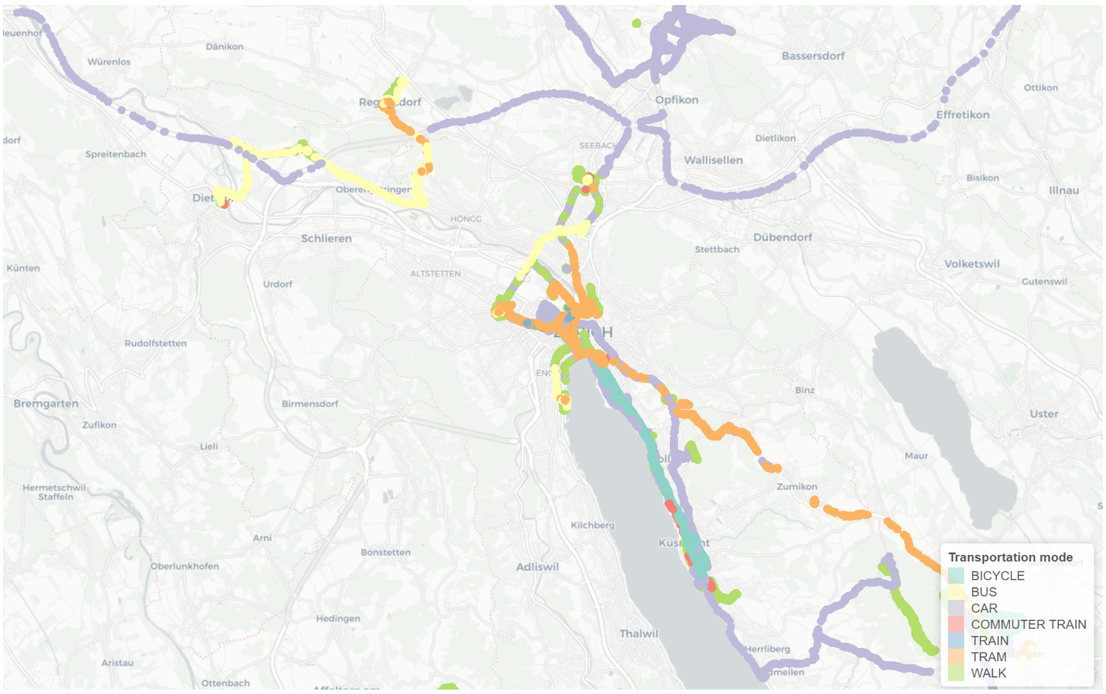

Figure 3.10: Close-up of all trajectories from the selected MOASIS dataset collected in the urban environment of Zurich.

In Figure 3.11, the distribution of the amount of data per minute is shown in comparing the benchmark dataset in light blue, and the MOASIS data in light green. The relative distribution of the two datasets is similar, with walking occurring most often, followed by the use of car. Walking is the primary transportation mode, as it always occurs at the beginning and end of the use of other transportation modes. Additionally, walking outdoors is generally better recorded by the GPS than the use of a transportation mode where the individual is out of reach for the satellites. As for the MOASIS data used, car is the second most often used mode of transportation. Many of the selected individuals live in rural areas, where the use of cars is often necessary. For testing the classifier, it is not ideal to have such little data for the public transportation modes.

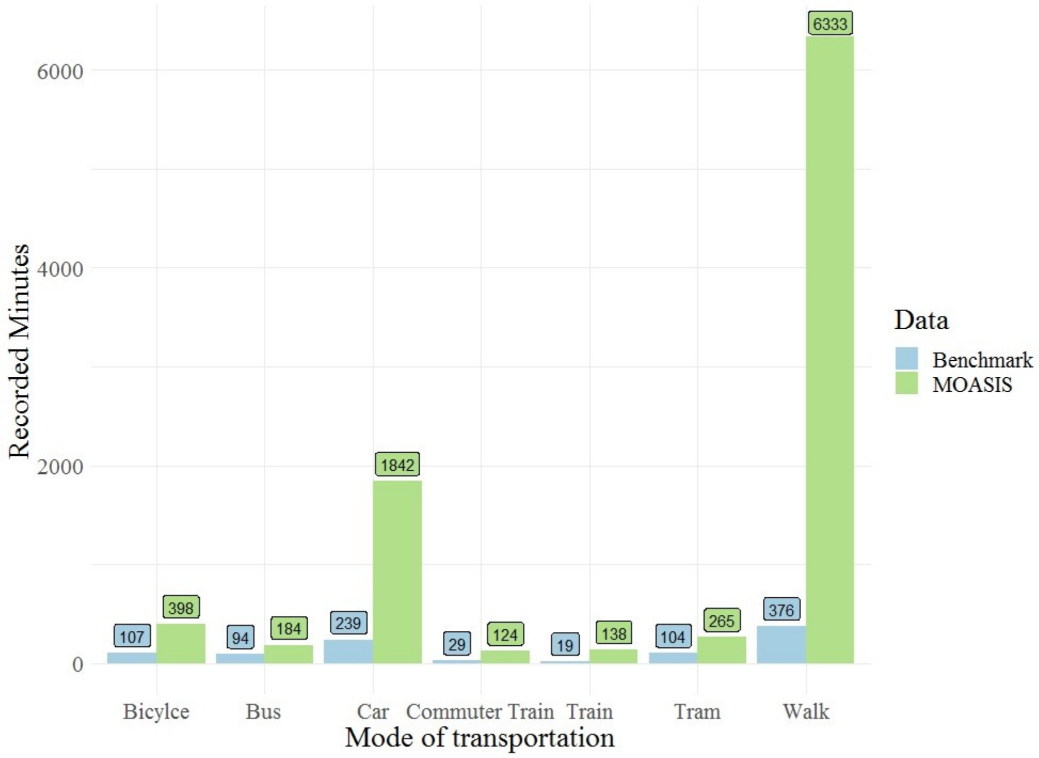

Figure 3.11: Recorded minutes per transportation mode for the benchmark dataset and selected MOASIS data.

An initial concern towards using a classifier trained with data collected by people from a different age group is the potential difference in the characteristics of the walking segments. All other transportation modes, especially passive ones, are generally independent of the age or fitness of the individual recording it. Figure 3.12 visualises the median values for speed, where no significant difference between the two datasets is visible. Opposed to the intuitive reasoning, data collected within the MOASIS study by older individuals has slightly higher median speed values.

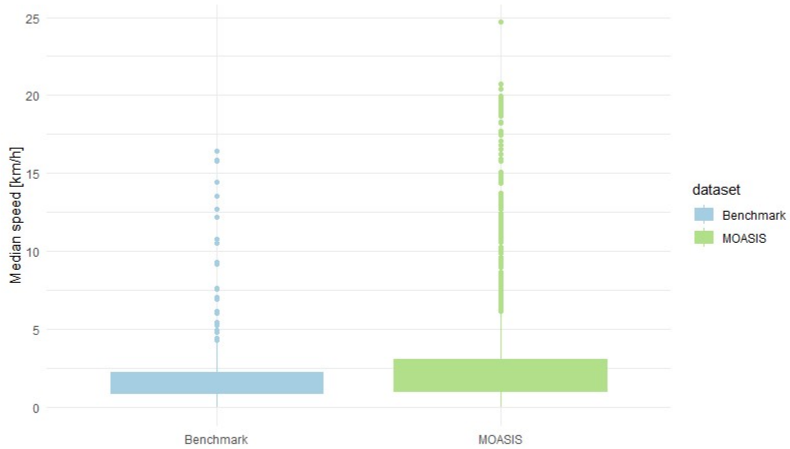

Figure 3.12: Median speed for benchmark and MOASIS data.

4 Methods

The goal of this thesis is the classification of unlabelled individual mobility data into the seven transportation modes defined by the benchmark dataset. The idea is to analyse the difference in classification accuracy depending on the input sensor, which is essential towards the MOASIS dataset that does not have full coverage by both sensors. To achieve this analysis, the entire classification process is done using the previously introduced labelled benchmark dataset to train the classifier. The classification process consists of multiple steps following the procedure described by Reddy et al. (2010), consisting of pre-processing, segmentation, feature selection, classification and validation. The procedure of each of these steps illustrated in Figure 4.1 is described in the following section. The complete analysis is performed using R version 3.4.1 (R Core Team, 2017). An overview over the different scripts with their most important functions can be seen in Table 8.5 (see Appendix 2).

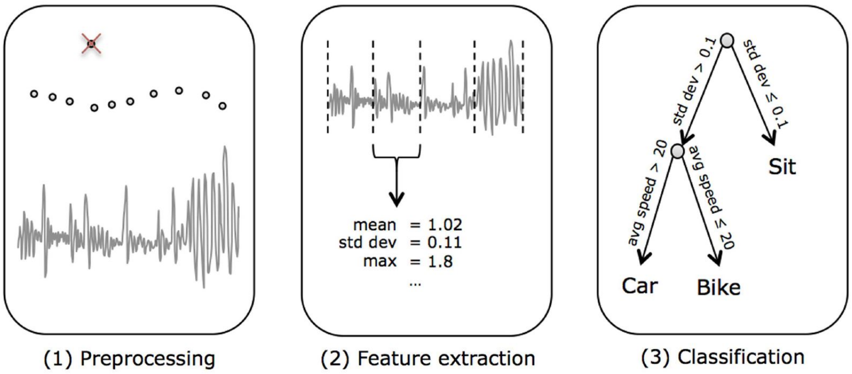

Figure 4.1: Visual representation of the neccessary steps for classification (Ellis et al., 2014).

4.1 Pre-processing

The benchmark data does not require pre-processing, and the data quality is high, though gap removal has been performed to ensure that the entire benchmark dataset has complete coverage by both sensors (Reddy et al., 2010). Temporal completeness of data is ensured using a gap-finding function, which determines how large the time gap is between two following records. The ensuring of temporal completeness is essential for testing the segmentation methods, as they only work on temporally coherent trajectories. For the entire thesis, the combination of GPS and accelerometer data always involves a loss in resolution for the accelerometer data, as only entries are used where both sensors have a record at a specific time.

Data collected within the MOASIS study needs additional pre-processing compared to the benchmark data. In addition to the removal of invalid points and incomplete sensor coverage, this consists of the removal of long stationary periods and transportation modes different from the seven selected ones. The data is cleaned during the visual inspection and manual labelling of transportation described in Section 3.3. That step ascertains the use of valid data for the classification of mobility. After the first data-cleaning, the resolution is assessed and the extraction of data with high resolution performed.

4.2 Segmentation

During segmentation, a trajectory is partitioned into a minimal number of segments, while satisfying the spatio-temporal criteria required for its homogeneity (Buchin et al., 2011). As the recorded trajectories could have been undertaken using different transportation modes, the goal of a segmentation method is to derive single-mode segments. The points where transportation mode changes are defined as transition points, with the first and last point of a segment being the transition point into another segment (Biljecki et al., 2013).
Generally, there are two different segmentation methods described in the literature, point-based methods and segmentation-based methods. Point-based methods are mostly used when determining transportation modes in real-time. An algorithm analyses the different features of a point in space and assigns a suited transportation mode to that point close to real-time. Segmentation-based methods are used to classify already collected data into transportation modes. This is done in detecting where the transportation mode changes and finding those changing points using heuristic rules. In the post-processing stages, a transportation mode is assigned to the entire segment (Prelipcean et al., 2016). For this thesis, a segmentation-based approach has been chosen because the data collection is completed.

Segmentation can be done based on simple rules, summarising an individual's travel behaviour. Such behaviour is people walking between different modes of transport, or the improbability of a specific combination of mobility modes (Guo et al., 2012). Biljecki et al. (2013) assume that a stop is necessary while changing transportation mode and set the transition points where the stopping time is over a certain threshold. This method could lead to a large over-segmentation as a stop does not always imply a change in mode. The idea behind it is not to miss quick transitions between transportation modes. Besides, in post processing, two consecutive segments that are assigned the same transportation mode can be merged. Even with this approach, it is challenging to capture swift transitions. Further, segmentation methods using more variables, including recorded or calculated features such as velocity, change point detection or broader spatio-temporal features, such as location, heading and speed, showing good results (Guo et al., 2012).
In addition to the raw records of speed and acceleration, further input variables for the segmentation process are suggested by the literature. For the GPS sensor, only the records of speed are used to create features characterising different modes of transportation. While the 3D accelerometer signal has three measured accelerations in the x-, y- and z-direction, its orientation and influence are unknown; thus an acceleration signal summarising the three axes is proposed. The device may have flipped or rotated during the data collection; thus the total acceleration is calculated $TotalAcceleration = \sqrt{x^2 + y^2 + z^2}$ capturing the overall magnitude of the acceleration, which proves to be the simplest and most effective solution to remove the influence of orientation (Lu et al., 2018).

The segmentation methods will be implemented on the benchmark dataset to determine adequate thresholds and validate their performance with the recorded transition points. Eventually, the most accurate segmentation method is applied to selected data from the MOASIS dataset and different thresholds are tested after analysing values from the literature. For results that look promising, the difference between the calculated and real transition points is quantified, with an accepted time variation of one minute. Besides, the amount of large over- respective under segmentation is taken into account for determining the best segmentation thresholds. The analysis is loosely based on the validation procedure described by Prelipcean et al. (2016) to find the agreement between the computed transition points and the ground truth. After the initial analysis, a strong over-segmentation is visible. To reduce this over-

segmentation, an approach described by Zheng et al. (2010) has been applied, in defining the minimum length of a segment as one minute.
Based on different values, such as the correctly calculated transition points, a quantitative way of evaluating the segmentation results is applied. It must be stated that generally, scientists that work on trajectory segmentation report the outcome of the used algorithms with traditional metrics that fail to cover all dimensions of error in dealing with continuous intervals and error propagation (Prelipcean et al., 2016).
The following sections introduce the three tested segmentation methods.

4.2.1 Change-point based segmentation method

The change-point based segmentation is based on the simple assumption that a person must walk while changing transportation mode, even if only for a short time. Beginning- and endpoints of a walking segment are thus detected as transition points using the assumption that while changing transportation mode, the velocity and acceleration drop (Zheng et al., 2010). This method identifies the beginning and end of a walking segment, in analysing and selecting the characteristic features of walking, such as low speed (fig. 4.2 (a)) and high frequency of change in total acceleration (fig. 4.2 (b)) captured by the two sensors.

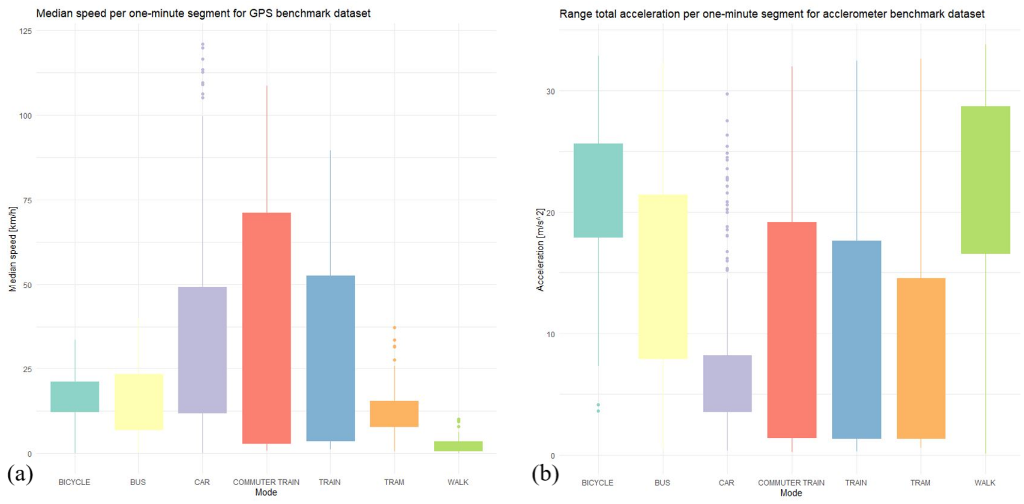

Figure 4.2: Figures illustrating the differentiating characteristics of walking segments compared to the other transportation modes. In (a) the median speed for the GPS segments is shown and in (b) the range of total acceleration per transportation mode.

Following the illustrated characteristics, the thresholds that are used to set a transition point have been determined by analysing the benchmark data and from the literature. A loose upper bound on speed should be selected, the mean of walking is in the range 6 - 12km/h and the maximum walking speed is of 11 - 16km/h. For the accelerometer records, the thresholds are deviated after analysing the walking segments of the benchmark data (Feng & Timmermans, 2013). It was found that the standard deviation and the difference between the maximum and minimum acceleration distinguish walking segments best from other transportation modes. They capture the higher frequency of signal acceleration and deceleration while walking.
A function is applied over a selected time window retrieving simple statistical values, such as the mean,

median, maximum, minimum and standard deviation for both speed and total acceleration. Additionally, as suggested by Shafique & Hato (2014), the range, consisting in the difference between the minimum and maximum acceleration, showing the spread of the acceleration and deceleration in a segment, is calculated. The number of entries used for the calculation of the statistical values is defined by the window size, which influences the classification with bigger windows leading to a more reliable prediction as more information is collected (Jahangiri & Rakha, 2014). Whereas too small of a window is prone to be influenced by noise, too big of a window may smooth the data (Yu et al., 2014). For the segmentation of transportation modes, the window must be big enough to capture periodic signals of activities such as walking.
The segmentation method has been implemented manually, using the *rollApply()* function from the *rowr-package*; simple statistical measures are calculated for a defined window size (Varicchio, 2016). For a segment to be reliably classified, a window size of one minute is required (Huss et al., 2014). Different window sizes are tested, with lengths of 3, 2, 1 and 0.5 minutes. In comparison, bigger window sizes perform worse in recognising transition points, whereas the difference between a one minute and 30-second window is minimal. The final window size is set to one minute, which is small enough to depict the characteristics of a transportation mode without over-smoothing the data. For the GPS sensor with a frequency of 0.5 Hz a one-minute window contains 30 recordings and for the accelerometer signal with a frequency of 25 Hz, 1500 recordings describe one minute. In the algorithm the window to calculate the statistical values for the data is aligned to the left, this means the calculation takes into account the points after the point of interest.
For the selection of the speed and acceleration features the starting point was informed by prior literature, different speed acceleration values were tested. Speed thresholds yielding the best segmentation results are a maximum velocity of below 10 km/h or a median velocity of 6-12km/h. Accelerometer thresholds giving the most accurate segmentation were found to be a standard deviation of over 2 m/s^2 or a range of acceleration bigger than 20 m/s^2. When using the data from both sensors, the threshold values are combined. The values for detecting walking segments are summarised in Table 4.1.

Table 4.1: Selected threshold values for the change-point based segmentation method.

Sensor	Thresholds	Window size (seconds)
GPS	Max <10 km/h ∣ Median 6-12 km/h	30
Accelerometer	Standard Deviation $\geq$ 2 m/s^2 ∣ Range >20 m/s^2	1500
GPS & accelerometer	Combination	30

4.2.2 Transition-oriented segmentation method

The transition-oriented segment detection finds features of the GPS and accelerometer measurements having differences between two consecutive recorded locations exceeding a specific threshold value (Prelipcean et al., 2016). The recorded characteristics are analysed, and an attribute function is applied to describe a value at every point over a particular time window. Afterwards, a criterion is defined to analyse the behaviour within a trajectory; for example, speed values cannot be derived for a single point within a window. This application leads to the similarity of the segmentation attributes within each segment (Buchin et al., 2011).
Using the assumption from Zheng et al. (2010) introduced in the change-point based segmentation method, the most distinguishable transportation mode out of the labelled ones is walking. This distinction generally means that the features of walking segments are analysed and have been used as thresholds for the segmentation, but more because of the large contrasts to other transportation modes.

This segmentation method has been implemented using the available R-package *segclust2d* (Patin et al., 2019). Initially written for ecological segmentation, it can be applied to different types of time-series, segmenting multivariate time-series based on homogenous behaviour. The algorithm partitions a time-series into segments containing similar speed and accelerometer measurements. Segmentation is then performed using a dynamic programming algorithm, finding the most accurate segmentation based on the inputs. The exact maximum likelihood solution is reached with a complexity of O (n^2*K), with n being the positions where the signal can be cut and K the number of segments. In using the dynamic programming algorithm, the complexity is reduced compared to a force brute algorithm (Patin et al., 2018). The trajectory is sectioned into segments based on the assumption that the mean and variance of a segment's recorded variable can only have a limited number of values within a segment (Picard et al., 2007). The algorithm selects the best number of segments based on a criterion developed by Lavielle (1999); the criterion is based on the value of the second derivative of the penalised likelihood. For an unknown number of transition points, the best number is selected in minimising a penalised contrast function. A likelihood value is calculated and associated with all the possible number of segmentations (1:kmax). Moreover, the log-likelihood of the best number of segments can be plotted and where the curve has a precise breakpoint the Lavielle-selected optimum is found. An example of this plot is shown in Figure 4.3, where the log-likelihood for a specific trajectory is displayed. The appropriate number of segments is determined to be four, following the bend in the curve.

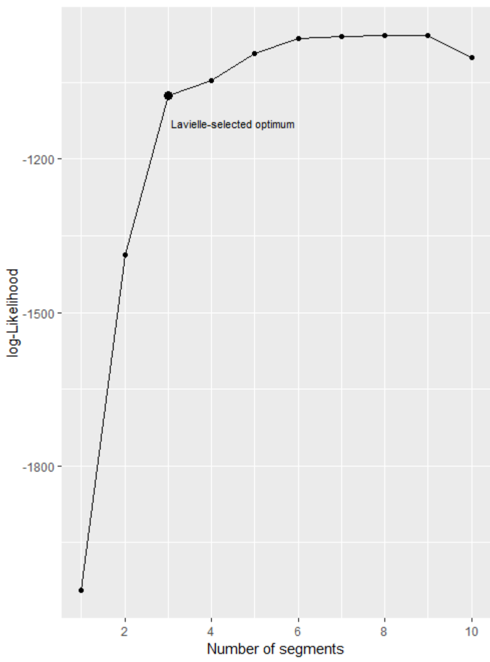

Figure 4.3: Plot of the log-likelihood to select the optimal number of segments for a trajectory with a Kmax of 10.

An advantage of the algorithm is the requirement of only a few inputs, such as *Lmin* describing the minimum length that a segment must have and *Kmax* determining the maximum number of segments allowed to segment a trajectory. A high *Kmax* can slow the calculation and lead to over-segmentation. *Seg.var* is the input variable that is used by the algorithm in order to segment the trajectory. With the benchmark data, different *Kmax* values are tested with the values of 10, 15, 20, 25 and 30. As *Kmax* determines the maximum number of segments, it influences the sensitivity of the segmentation. For the benchmark data, the value of 25 proves to be a proper equilibrium between minimising transition point errors and controlling over-segmentation. For the value of *Lmin*, following the explanation given in Section 3.2 about the minimal length of a segment, the number of entries describing one minute of data for the corresponding sensor is selected. Table 4.2 shows the thresholds for the segmentation with the transition-oriented method.

Table 4.2: Selected input values for the transition-oriented segmentation method.

Sensor	Kmax	Lmin
GPS	25	30
Accelerometer	25	1500
GPS & accelerometer	25	30

4.2.3 Fixed time window segmentation method

Most segmentation approaches discussed in the literature are based on some simple assumptions of travel behaviour. As they may not correctly represent real-life behaviour, an additional segmentation method is tested, which is based purely on the temporal-segmentation of a trajectory. Instead of the segmentation of the trajectory into segments describing one transportation mode, the trajectory is segmented into uniform temporal segments, which are then individually classified. Data is split into the often suggested time interval of one minute and features thereof are extracted (Ellis et al., 2014).
This segmentation method has some advantages: It can be used on data that has temporal gaps, as only one minute of data is necessary for classification, making it more suitable towards the application on the MOASIS data. Additionally, for the transition-oriented and change-point based segmentation methods, there is the problem of error propagation. If a transition point has been determined wrong along the trajectory, the features calculated from that segment are influenced by different modes of transportation, making the correct classification even more difficult. In calculating the features over a fixed window size, this error propagation is minimised.

4.3 Feature Selection

The selection of features is essential in reducing the dimensionality and noise of the data and identifying the essential classifying predictors (Jahangiri & Rakha, 2015). Generally, features can be categorised into time and frequency domain. Time-domain features characterise information within the time-varying signal, such as the speed and acceleration measurement along the trajectory. These features consist of raw speed or acceleration data as input features. The feature domain consists of computationally more demanding features, due to the need for an additional processing step transforming the data from the time domain. An example of such a feature derived from the accelerometer signal is the peak frequency of the power spectral density (Nikolic & Bierlaire, 2017).
Towards the research questions of this thesis, features that have few correlations with velocity are of particular interest, as they exclude the influence of traffic and show characteristics of vehicles. Such features can be direction change rates, velocity change rates and stop rates. Direction and velocity change rates help to distinguish between different types of passive transportation, and the stop rate helps distinguish driving in a car versus in a means of public transport (Zheng et al., 2010).
The type and number of useful features are determined after the feature analysis of the segmented trajectories. In literature, the amount and selection of features used for the classification varies greatly. Additional features that come with the segmentation can be used, such as trajectory length, stop rate,

amplitude and change frequency within a segment (Xiao et al., 2017).
In Table 4.3, all the features that were selected to characterise the segments with the different sensors and their sources are listed. For the GPS data, 20 different features are derived from speed and location, whereas for the acceleration data a total of 59 features are computed, consisting of 15 different features applied on three axes of the 3D accelerometer and the total acceleration.

Table 4.3: List of all the features computed over the segments of the trajectories.

Sensor	Computed features
GPS	Minimum, Maximum, Standard Deviation, Variance (Kaghyan & Sarukhanyan, 2013) 95^{th} percentile, Mean, Median (Biljecki et al., 2013) Interquartile range, 20^{th} percentile, 80^{th} percentile (Martin et al., 2017) Root Mean Square (Pires et al., 2017) Skewness, Kurtosis ((Pires et al., 2017); (Xiao et al., 2017)) Range of speed, Heading change (sum, max, min, range), Stop rate, Coefficient of variation (Xiao et al., 2017)
Accelerometer	Maximum, Minimum (Kaghyan & Sarukhanyan, 2013) Mean, Median, Standard Deviation, Variance (Feng & Timmermans, 2013) Number of Zero crossings, Peak-to-peak value (delta acceleration), Root Mean Square (Pires et al., 2017) Kurtosis, Skewness ((Pires et al., 2017); (Xiao et al., 2017)) Interquartile range (Jahangiri & Rakha, 2015) 20^{th} percentile, 80^{th} percentile, 95^{th} percentile

Features describing the characteristic behaviour of the modes of transportation are applied both on the data from the GPS and accelerometer sensors, which consists of measures of speed and acceleration in the three axes and the calculated total acceleration. Different granularity features are suggested by Hemminki et al. (2013), such as frame-based features, peak features and segmentation features. Frame-based features containing statistical features described by Xiao et al. (2017) are listed in the following:

- Minimum, maximum speed/acceleration, describing the extreme values in a segment
- Mean of speed/acceleration, describing the overall tendency of the segment
- Median of speed/acceleration, describing the value most often occurring in a segment
- Standard deviation and variance of speed/acceleration, describing the distribution of values within a segment
- Range of speed/acceleration, where the minimum value of a segment is subtracted from the maximal value of a segment showing the distribution of values
- Different percentiles of speed/acceleration, for this thesis the 95^{th}, 80^{th} and 20^{th} percentiles are used, describing the range of extreme values in a segment
- Interquartile range of speed/acceleration, describing the difference between the high and low quartiles
- Root Mean Square (RMS) of speed/acceleration, which is a more generalised version of the mean

The selected peak features described by Xiao et al. (2017) are the following:

- Kurtosis of speed/acceleration, where the tailedness of the probability distribution is determined and compared to the normal distribution. A positive kurtosis value describes a pointy curve, and a negative kurtosis a flat curve.
- The skewness of speed/acceleration used to analyse the distribution of the variable by measuring the asymmetry of the data compared to the normal distribution. A positive skewness describes a tail drag to the right and a negative to the left. The higher the value, the higher the asymmetry to

the normal distribution.

- The coefficient of speed variation, which measures the dispersion of the probability distribution in a standardised way, dividing the standard deviation by the average of the data.

Different segmentation features derived and adapted from Zheng et al. (2010) are explained below, describing features that are also useful to describe transportation modes. They state that the heading change rate and stop rate have few correlations with velocity, removing the influence of traffic conditions:

- Heading change rate for GPS points, which measures the turning angles occurring during a segment. No absolute value is used, but rather the sum, maximum, minimum and range of the turning angles, with the intent of differentiating further between modes of transportation moving in a steady direction and more dynamically changing (e.g. motorised and non-motorised transportation modes). The feature vector is calculated with the *TrajAngles()* function from the *trajr* package (McLean & Skowron Volponi, 2018). The information about the turning in a segment is tested to distinguish between public and private transportation modes. Especially in an urban environment, buses, trains and trams have tendentially less sharp turns in their path than cars.
- Stop rate of speed, counting how often in a segment the speed of an individual drops under the defined threshold of 0.05 km/h. This feature should differentiate between public and private transportation. Differently from Zheng et al. (2010), the number of stops is counted, without relating it to the distance. In Figure 4.4, the varied stop behaviours of the transportation modes are illustrated.

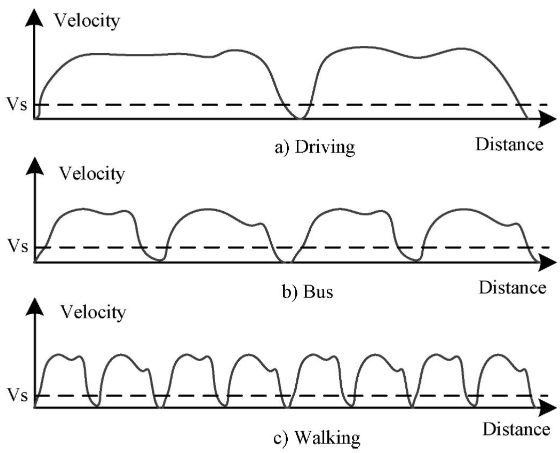

Figure 4.4: Stop rate for different transportation modes (Zheng et al., 2010).

- The zero-crossing of each accelerometer signal axis counts the number of times that the signal crosses the line of no acceleration. As the total acceleration is always positive, no zero crossings occur for that variable. Especially active modes of transportation with a periodical behaviour can be characterized by this feature, as it can be seen in Figure 4.5, where the zero crossings per transportation modes are summarised for data from the benchmark dataset (Pires et al., 2017).

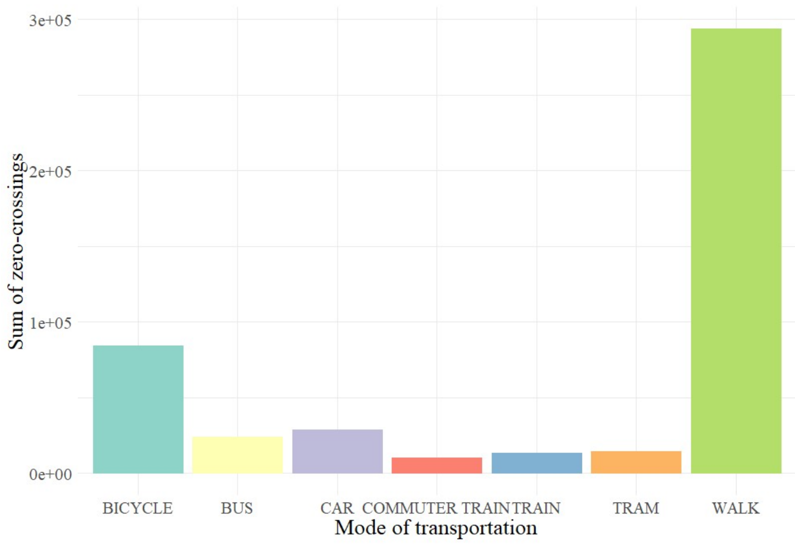

Figure 4.5: Sum of zero-crossings over all axes per transportation mode for the benchmark dataset.

The idea of this feature selection is to identify features that are not too complicated but with the specific goal in mind to classify the challenging modes of transportation. In the literature, some suggest that good results can be achieved in using only features such as the 95^{th} percentile of speed, acceleration as well as deceleration for the transportation mode detection with GPS Huss et al. (2014), where others state the use of average values and standard deviation to describe patterns of speed and acceleration. To measure speed patterns, it is suggested to use the average, maximum, and to include the non-moving time duration over a defined time frame (Feng & Timmermans, 2013). Ellis et al. (2014) suggest that the most important feature is the standard deviation of the acceleration, capturing the signal variability. Based on the different suggestions from literature, all the above-described features are generated for the segments and used to train the classifiers. After the classification, the ranking of the features is analysed and can be used for future classifications.

4.4 Classification

The classification is implemented by using data from the benchmark dataset for training classifiers. The first supervised classifications within this thesis are performed using this labelled dataset containing the seven transportation modes (Isler, 2018). Following the poor results occurring from the training of the classifier with segmentation by labels, the segmentation by a fixed window size is determined to be the better method.

After the first poor results of the classification, the hierarchical classification is also implemented. It starts more general and leads subsequently to the classification into each transportation mode. This approach to classifying transportation modes proposed by Hemminki et al. (2013) has been introduced in Section 2.2.2. The four-stage hierarchical classifier classifies the data into the main categories first (e.g. motorised or non-motorised) and is further divided into the individual transportation modes in the following steps.

Both the training and use of one classifier to either classify the dataset into the final classes or subdivide it into additional sub-categories for classification is described by one of the coloured boxes in Figure 4.6.

The four classifiers trained within this multistep classification process are the following:

1. Blue box: Classification between non-motorised and motorised transportation modes. The active transportation consists of walking and bicycling, where the class motorised transportation modes consists of car, bus, tram, train and commuter train. This distinction is especially interesting in a health context, documenting the positive influence of physical activity performed (Fillekes et al., 2019).
2. Red box: Classification between the two non-motorised transportation modes walking and bicycling.
3. Green box: Classification of motorised transportation modes in private and public transportation. Private transportation consists of the transportation mode car and public transportation of the remaining four modes.
4. Yellow box: Final classification of public transportation modes into the respective four classes, which are commuter train, train, bus, tram.

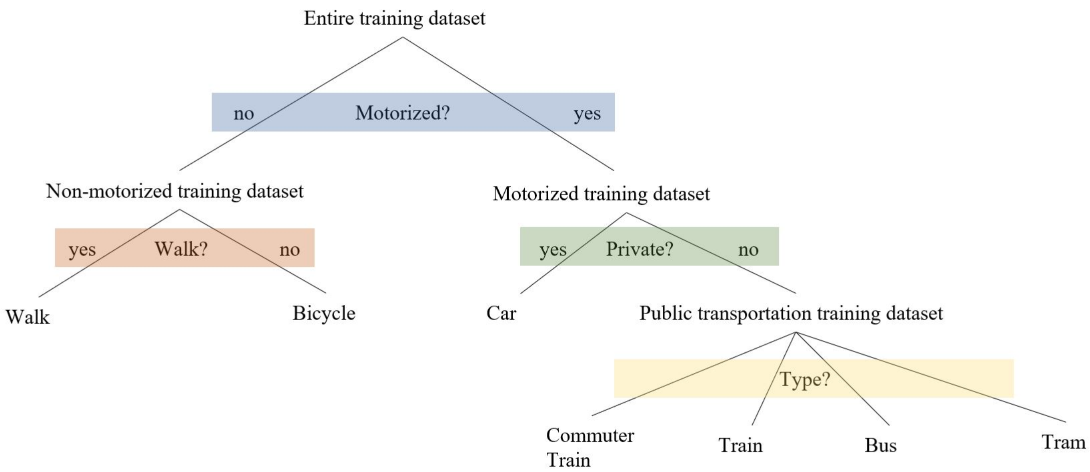

Figure 4.6: Proposed hierarchical classification approach, using different classifiers. Initially the distinction between active and passive transportation is performed. For the active transportation modes the classification into the mode bicycle and walk is made. For the passive transportation modes an additional classifier distinguishes between private and public transportation, and finally the four public transportation modes are classified.

After the training data analysis, the imbalance of the dataset regarding the distribution of data per transportation mode is striking. In Figure 4.7, the sum of entries per transportation mode is shown after they have either been segmented by labels or by the fixed window size. Walking occurs most often in the dataset, which is typical for transportation mode detection, as seen in other studies (Shafique & Hato, 2014).

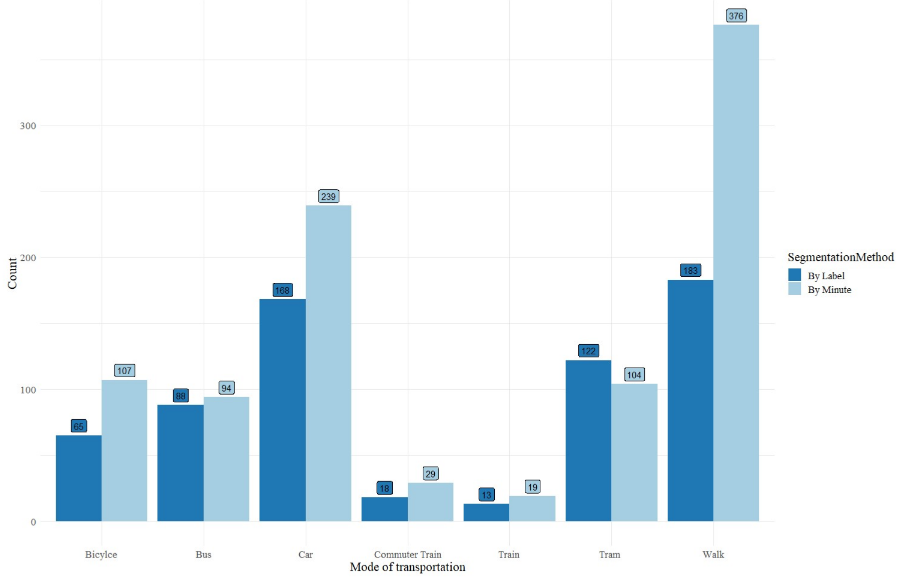

Figure 4.7: Segments per mode of transportation in the benchmark dataset segmented either by label or by minute.

This strong imbalance is considered when training the classifiers, as a performance bias can be experienced following this data distribution. Classifiers that are trained with such a biased training dataset lead to a high classification accuracy for the most frequent class and weak for the others (Bhowan et al., 2013). To deal with this imbalance, the data has been "up-sampled", meaning that the minority classes are randomly replicated to match the size of the majority class.

The following classification methods are all trained using 70% of the data for training and 30% for testing (Jahangiri & Rakha, 2015). As the number of segments is small, 10-fold cross-validation is applied for the classifier training, which is the state-of-the-art example used in data mining and processing systems (Kaghyan & Sarukhanyan, 2013). In the 10-fold cross-validation, the training data is further sampled randomly into ten subsamples. Where nine of these samples are used to train the classifier, one sample is used to test the trained classifier. In repeating this process ten times, each of the subsamples is used once to test the classifier and nine times to train a classifier. The results of these ten trained classifiers are finally averaged and result in one classifier (Stenneth et al., 2011).

As discussed in Chapter 2, different classifiers can be used for the detection of transportation modes. Three statistically robust machine learning methods showing good results in the literature are the Support Vector Machine, the Random Forest algorithm and a Neural Network (Prelipcean et al., 2016). Different software packages exist to implement the machine learning methods, using them as black-box models for the data classification (Ellis et al., 2014).
For the training and testing of the following machine learning methods the package *caret* (classification and regression training) has been used, implementing functions to streamline the process of classification (Kuhn, 2019). Each classifier is trained three times, resulting in not only the determination of the performance of the different classifiers but also of the different sensor combinations.
Drawbacks in using machine learning are not only the necessity of plenty, good and labelled training data but also the difficulty of interpreting rules learned by the algorithms (Prelipcean et al., 2016). The

used classification algorithms are ready-to-use; only the input data and some parameters are defined. The rest of the classifier training process happens within the algorithm and is not visible or influenced by the user. This opacity makes it challenging to understand occurring problems in the classification process and improve it for the specific data.
In the following sub-chapters, each of the tested classification methods is shortly introduced.

4.4.1 Support Vector Machine

The Support Vector Machine (SVM) is a large margin classifier; thus, it determines the best possible decision boundary between the different classes. The ideal case assumes that a linear hyperplane can separate two classes by the most significant margin (Lin et al., 2002). In the simplest version, the SVM displays all data point on a plane, separating the different classes by an optimal separation gap. New data points are classified depending on their position on the plane (Bedogni et al., 2012). In a two-dimensional space, a hyperplane consists of a line, whereas for the three-dimensional space, it consists of a plane. If the data is well separable by a hyperplane, there are an infinite number of hyperplanes, shifting the position slightly, without touching the observations. Even if a separating hyperplane would be possible, it might not always be desirable as it can be highly sensitive to change in observations. If there is a small margin between the hyperplane and observations, the confidence of correct classification is reduced. The ideal is to have many observations with a high margin; this can be achieved in allowing some observations to be on the incorrect side of the margin and even of the hyperplane (James et al., 2013). Compared to other classifiers, such as traditional Neural Networks, SVM has a simple structure and processes faster. A problem for the accuracy of the results for the SVM classification is an unbalanced dataset (Xia et al., 2014).

In order to construct the SVM model, a Gaussian kernel with complete model selection needs to be used, as well as feature scaling and examination applied. The kernels describe features that are passed to the classifier, based on the attributes of the data (Jahangiri and Rakha, 2015). Such kernels are used in order to deal with non-linear boundaries between classes, enabling the enlarging of the feature space. A kernel is a function that can quantify the similarity of different observations. Different types of kernels are used, such as:

- The linear kernel, which quantifies the similarity of observations using Pearson's correlation.
- The polynomial kernel, which can fit a support vector classifier to a higher dimensional space involving polynomials of degree d. This kernel is generally known as the SVM.
- The radial kernel, which is another non-linear kernel, with a local behaviour as it uses only close training observations, which have an effect on the class label of a test observation.

The use of such kernels compared to the enlargement of the feature space saves computational time (James et al., 2013). Figure 4.8 illustrates the different behaviour shown by different kernels. In Figure 4.8 (a), a polynomial kernel has been applied to non-linear data, enabling the partition of the two classes. In Figure 4.8 (b), a radial kernel has been applied, showing a different behaviour due to the interaction with local points around it.

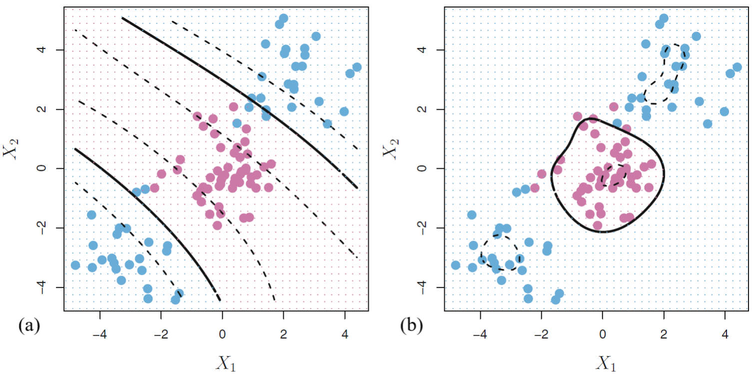

Figure 4.8: Illustration of the behaviour of different kernels in dividing the dataset into two classes. On the left (a) a polynomial kernel has been applied onto non-linear data and on the right (b), a radial kernel is applied (James et al., 2013).

The selected kernel for the SVM applied in this thesis is a Radial Basis Function Kernel, where the features of the data are mapped to a higher-dimensional space. In comparison to the selected radial kernel, the linear kernel cannot handle non-linear attributes and is not suited. The polynomial kernel, with more hyperparameters, is more complicated, but even with the higher complexity, the results of classification are not improved (Hsu et al., 2010).
For the selected method, the input parameters to be determined for the classifier are cost and sigma. The parameter cost (C) defines the margin of the severity of violations of the hyperplane. If C is equal to zero, then no observations can violate the margin, meaning no observations are tolerated on the wrong side of the hyperplane. With the small value for C, the classifier is perfectly fit for the data, describing low bias and high variance, where a larger C allows more violations, fitting the data less strictly. The observations lying on the margin are called support vectors. With a larger C, more observations lie on the margin, increasing the number of support vectors defining the hyperplane. The decision rule of a support vector classifier is determined by only a small subset of training observations, making it robust to the observations lying far away from the hyperplanes (James et al., 2013). Sigma is a parameter that defines how quickly the similarity metrics moves towards zero for the points that are further apart (Peixeiro, 2019).
For this thesis, the *train()* function from the *caret-package* is used, with the *method = "svmRadial"* (Kuhn, 2019). The values for the parameters were extracted from the *$bestTune* information of the trained classifier and are listed in Table 4.4.

Table 4.4: Final tuning parameters extracted from the trained classifiers for SVM with the different sensors.

Sensor	Sigma	Cost (C)
GPS	0.077	1
Accelerometer	0.015	128
GPS & accelerometer	0.009	16

Towards the application on selected data from the MOASIS study and the size of the training dataset, this model's qualities lie in resolving problems that involve small samples, nonlinearity, regression and classification of high-dimensional patterns. Also, it generalises better, is easier to implement and processes faster than other machine learning methods (Xia et al., 2014).

4.4.2 Random Forest Algorithm

Random forest (RF) is an algorithm adapted from the Decision Tree method, which searches for features on each classification step dividing the data into distinct branches (Brondeel et al., 2015). The leaves of the Decision Trees can be seen as the classes (here: transportation modes), and the branches as the differences in the features leading to the classes. During the training of the algorithm, such a Decision Tree is built, growing branches and conjunctions based on the features. When data from the testing dataset is inserted into the classifier the final classification of this point is achieved in moving through the branches of the Decision Tree, when a leaf is reached said class is assigned to the data point/feature vector (Ellis et al., 2014). An example of such a Decision Tree for the transportation mode data collected by the GPS sensor is illustrated in Figure 4.9. In the coloured boxes, it can be seen based on what features and which values the distinction between the different classes is made.

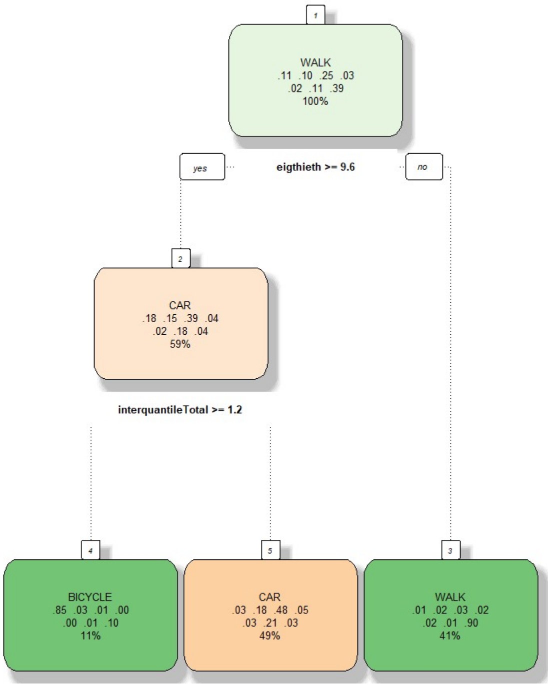

Figure 4.9: Example of a Decision Tree built with benchmark GPS data to distinguish between three different classes. At the bottom of the boxes the feature as values selected to distinguish the classes are shown (James et al., 2013).

The RF algorithm uses a large variety of variables to predict the transportation mode of a segment. In order to have better generalisability, it consists of the Decision Tree method, with an addition of two sources of randomness. Those sources of randomness are considering a part of the explanatory features for the definition of the branches, using a random subsample for building the trees. All the trees that were grown are summarised in a forest, in order to make a forest prediction, the transportation mode for each trip is obtained based on the summarised prediction of all trees (Brondeel et al., 2015). In doing this, strong predictors in the data do not lead to a similar signal in many trees, controlling the correlation between trees, which leads to a smaller variability but higher reliability (James et al., 2013).
As a method *"ranger"* from the *ranger-package* is selected, allowing a fast implementation of the RF algorithm (Wright and Ziegler, 2017). The parameters to define for the classification are *mtry, splitrule*

and *min.node size*, with their final values listed in Table 4.5. *Mtry* describes the number of candidate predictors, which are set at the square root of the number of features selected (Boulesteix et al., 2012). The *splitrule* is defined by the extra-tree algorithm, which splits nodes in choosing cut-points at random, using the whole learning sample to grow the trees (Geurts et al., 2006). The *minimum node size* describes the probability score for a transportation mode. It is set to the value 1, meaning that a node can consist of only one entry (Ellis et al., 2014). A small value is showing the few data entries for some classes. The number of trees has been set to 500.

Table 4.5: List of the selected parameters for the RF classifier by the sensor.

Sensor	Mtry	Splitrule	min.node size
GPS	4	Extra-trees	1
Accelerometer	7	Extra-trees	1
GPS & accelerometer	8	Extra-trees	1

4.4.3 Neural Network

Neural Network (NN) classifiers learn the underlying patterns of a dataset during multiple cycles of training. The NN consists of different layers of neurons connected by different weights and nonlinear functions within the neurons. During training, the classifier adjusts the connecting weights between the neurons in order to determine the most probable transportation mode (Byon et al., 2009). A NN contains one or multiple hidden layers between the input and output layers, as illustrated in Figure 4.10. In the hidden layer, the neurons take the output of the layer before and use it as an input into the next layer. The overall output of the neuron is a function of a weighted sum of these different layers (Mäenpää et al., 2017).

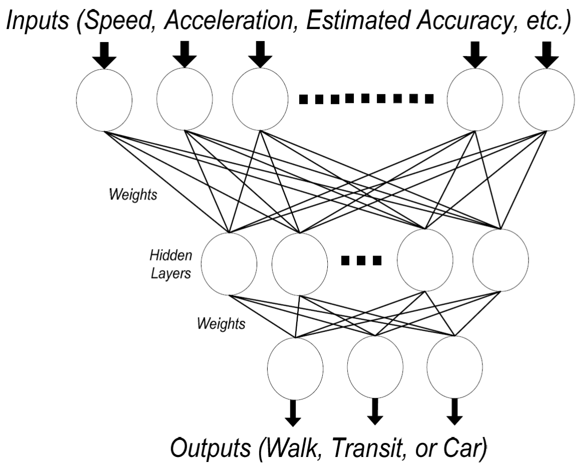

Figure 4.10: Structure of a NN classifier (Gonzalez et al., 2008).

The classification is performed using the Multilayer Perceptron (MLP) algorithm, which is the most frequently used NN for supervised learning (Byon et al., 2009). Due to the good suitability in pattern classification, it is used in different applications, with the number of output nodes representing different classes. The MLP is composed of different non-linearly interconnected nodes by different weights. After each node, the output is modified by the weight between the nodes and by a simple nonlinear transfer function, leading to the approximation of non-linear functions. MLP is a feed-forward NN due to the

defined direction of information processing. Information that has been outputted by a preceding node is scaled by the connecting weight and used as input for the next node. The full connectivity is described by the link of each node to the next and previous layer. During training, the weights are continuously adjusted by repeatedly inserting the training data into the model. The magnitude of the error is determined in analysing the predicted and actual output adjusting the weights. Unlike other methods, the MLP does not assume any distribution of the data (Gardner & Dorling, 1998).
The above-discussed classifier is applied in selecting the *method = "mlp"* (Kuhn, 2019). The tuning parameter for this classifier is size, which describes the number of units that are in a hidden layer. The parameters chosen by the classifier handling the data are extracted and listed in Table 4.6.

Table 4.6: Listing of the selected input parameter size for the multilayer perceptron applied to the benchmark data by the sensor.

Sensor	Size
GPS	19
Accelerometer	19
GPS & accelerometer	5

In addition to detecting patterns that humans and other algorithms might not see, the NN can generalise conclusions for data, without matching the training data completely. This generalisation is a possible strength in the transferability of testing a classifier with a different dataset than on which it has been trained. Besides, no assumptions need to be made about the data distribution, and no relative importance is defined between the measurements of the input data. Further, depending on the settings and data, the NN can automatically learn even subtle differences in transportation modes where other methods have more difficulties (Gonzalez et al., 2008).

4.5 Validation

In the final step, the validation of the different types of classifications is performed.
The displaying of the confusion matrix takes all relevant results into account, as it shows information about the correct modes and the results from the classifier (Xia et al., 2014). Large numbers show high accuracy in the diagonal and small numbers in the other boxes (Witten et al., 2017). Additionally, quantitative values to describe the results of the classifier are the precision, recall and F-score. Precision is the amount of correctly predicted transportation modes in percentage, calculated in dividing the true-positives by the sum of the false-positive and true-positive. Recall measures the proportion of right transportation modes that have been classified correctly calculated in dividing the true-positives by the sum of the true-positive and false-negative. The F-score is an indicator for accuracy, calculated by the harmonic mean of precision and recall (Ellis et al., 2014). In Figure 4.11, the terms precision and recall are shown with the help of a confusion matrix. The true positives are highlighted in yellow, and the false positive are highlighted in grey.

	Actual: Walk	Actual: Car	Precision
Pred: Walk	120	30	0.8
Pred: Car	16	84	0.84
Recall	0.88	0.74	

Figure 4.11: An exemplary confusion matrix is illustrated, showing true positives in yellow and false positives in grey. The calculation of precision (light blue) and recall (light green) is demonstrated with the matrix.

The above-discussed evaluation quantifiers are useful for a first overview, but the agreements occurring by chance are not taken into consideration. The kappa statistics measure considers the results of the random predictor and deducts it from the successes. This variable results in the accuracy for a perfect predictor, which omits the occurrence by chance (Witten et al., 2017).

5 Results

5.1 Segmentation

In this section, the results of the change-point based and transition-oriented segmentation methods are shown. To present the segmentation results, one trajectory containing multiple transportation modes is illustrated with the calculated transition points from each segmentation method (fig. 5.1 (a) & fig. 5.2 (a)). A red dashed line illustrates the labelled transition points and the calculated transition points are shown with a dark blue line. For such trajectories where the values for the different modes are so distinctive, transition points are generally detected well by both methods. In the following, the different characteristics of the results for the segmentation methods using different sensors are presented, using one trajectory of the benchmark dataset.

In Figure 5.1 (b & c), clear transitions between transportation modes are well detected in using the GPS sensor. For transportation modes with fluctuating speed, such as the car segment shown in red in the example trajectory, additional transition points are determined for the change-point based segmentation method (fig. 5.1 (b)). Real transition points are detected accurately, sometimes slightly before they occur. The transition-oriented segmentation method is based on an existing algorithm that partitions the trajectory in segments with similar speed measurements (fig. 5.1 (c)). This method does not unnecessarily define transition points in the car segment, as it considers it as a long enough segment to include the fluctuations and recognise it as the characteristics of that transportation mode. Especially focusing on the first transition point, it is set at the point of a most substantial change in speed. It is comprehensible that this is the case and not the recognition of the labelled transition point, which is slightly earlier and can describe the period of entering a train before it moves. So, the transition point is set at a comprehensible location, even though the individual labelling this trajectory changed transportation mode a bit earlier. This difference very well shows the difference in transition point detection between an algorithm and the individual labelling.

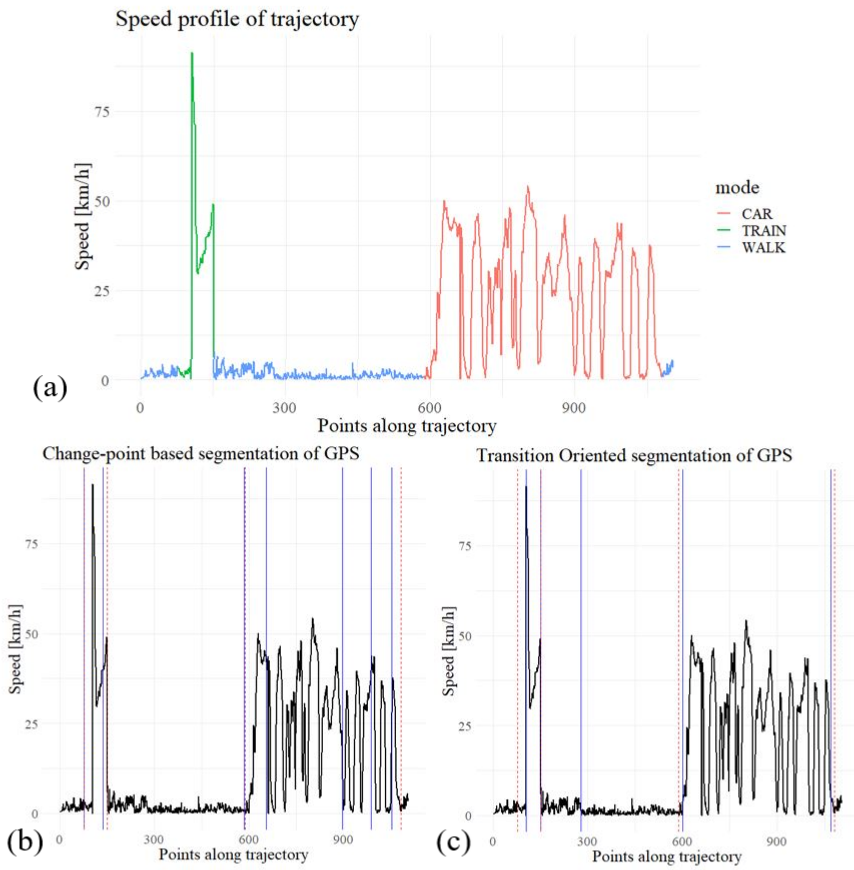

Figure 5.1: At the top the speed over time of the selected trajectory for the GPS sensor is shown (a). The line is coloured according to the transportation modes. At the bottom (b) shows the trajectory segmented by the change-point based method and in (c) the same trajectory segmented by the transition-oriented segmentation method. Red dotted lines show the labelled transition points and blue lines the calculated ones.

The segmentation results using the accelerometer are shown in Figure 5.2. It can be seen in Figure 5.2 (b), that for the change-point based segmentation, only three out of four transition points have been well detected. Due to the stronger fluctuation in total acceleration towards the end of the second walking segment, the transition point is set at the end of the walking segment with the highest fluctuation. Here a problem appears, due to the definition of the transition point too early, it influences the detection of the next transition point, which happens when the change of signal is not clear. The transition-oriented segmentation method detects the real transition points well, but due to the variability in the second walking segment, additional transition points are defined (fig. 5.2 (c)).

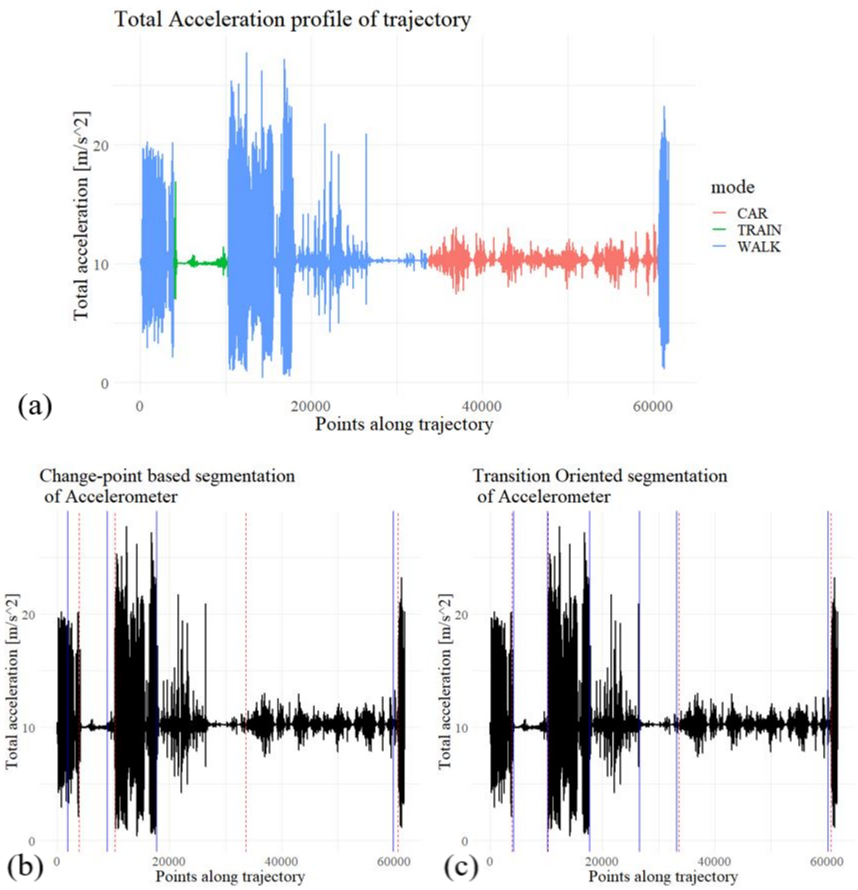

Figure 5.2: At the top the total acceleration over time of the selected trajectory for the accelerometer sensor is shown (a). The line is coloured according to the transportation modes. At the bottom (b) shows the trajectory segmented by the change-point based method and in (c) the same trajectory segmented by the transition-oriented segmentation method. Red dotted lines show the labelled transition points and blue lines the calculated ones.

For the segmentation illustrated in Figure 5.3, both sensors have been used to calculate the position of the transition points, with the display of speed on the y-axis. For the change-point based segmentation method (fig. 5.3 (a)), the use of both sensors for segmentation has improved the result. The over-segmentation of the car segment that occurred in using only speed as a variable is improved, and the definition of a transition point too early for the walking segment using the accelerometer data only is eliminated. For the transition-oriented segmentation method (fig. 5.3 (b)), the use of both sensors worsens the detection of transition points in additionally segmenting the walking segment. This segmentation can also be a chance to analyse different types of walking, as more subtle changes within a transportation mode are detected. In general, it must be stated that the over-segmentation is not necessarily bad, as the classifier can assign multiple consecutive segments to the same mode.

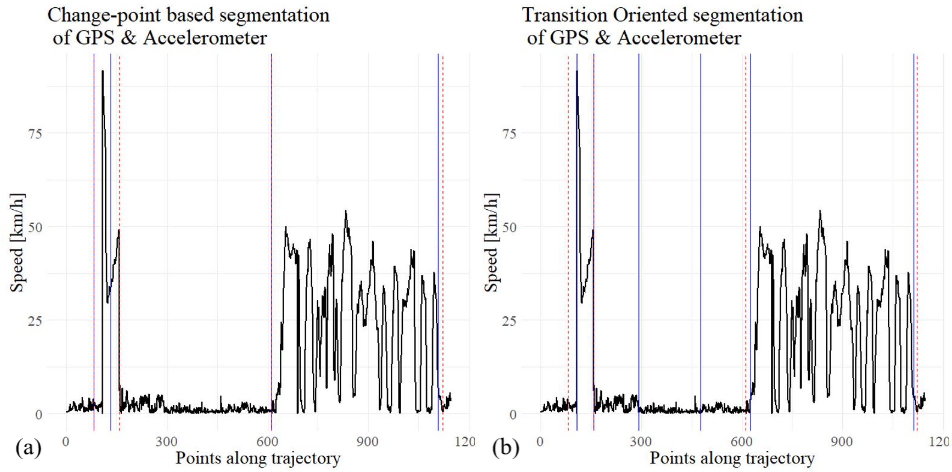

Figure 5.3: In (a), the segmentation of the trajectory using both GPS & accelerometer features by the change-point based segmentation is illustrated. In (b), the same trajectory has been segmented slightly differently by the transition-oriented segmentation method.

For the above-discussed trajectory, the results from both segmentation methods look promising. The change-point based segmentation method performs better in detecting transition points with the GPS signal (fig. 5.1 (b)), whereas the transition-oriented segmentation method shows better subdivision into the correct segments for the accelerometer data (fig. 5.2 (c)). Where both sensors are used, the change-point based segmentation recognises the transition points more accurately (fig. 5.3 (a)). As the ability to better detect the transition points depends on the characteristics of the recorded trajectory, the data strongly influences the performance of the segmentation methods. For both segmentation methods, the performance of the segmentation algorithms and calculation of the transition points is very slow, due to a large amount of data to be processed.
The algorithm for change-point segmentation has been developed within this thesis. After the segmentation, two significant disadvantages of this method have appeared:

- Periodical acceleration and deceleration, such as the typical behaviour of public transportation, are not recognised. Every time the defined thresholds of acceleration/deceleration are exceeded a new segment starts, calling for the improvement of the algorithm.
- Under-segmentation is an issue where transitions are not as clearly visible in the signal as in the example above. This helps prevent the segmentation of unimodal trajectories but reduces the correct detection of transition points for a bimodal trajectory severely.

In Figure 5.4, such a case of under-segmentation is illustrated. Due to the selection of the statistical variables, which are calculated over a time window, with the fluctuating behaviour of the speed records, the change signal is less clear than in Figure 5.3. For the change-point based segmentation method (fig. 5.4 (a)), this leads to an under-detection of transition points. Figure 5.4 (b), shows the transition-oriented segmentation, where the problem is the missing of real transition points. Especially short segments are difficult for the segmentation method to detect.

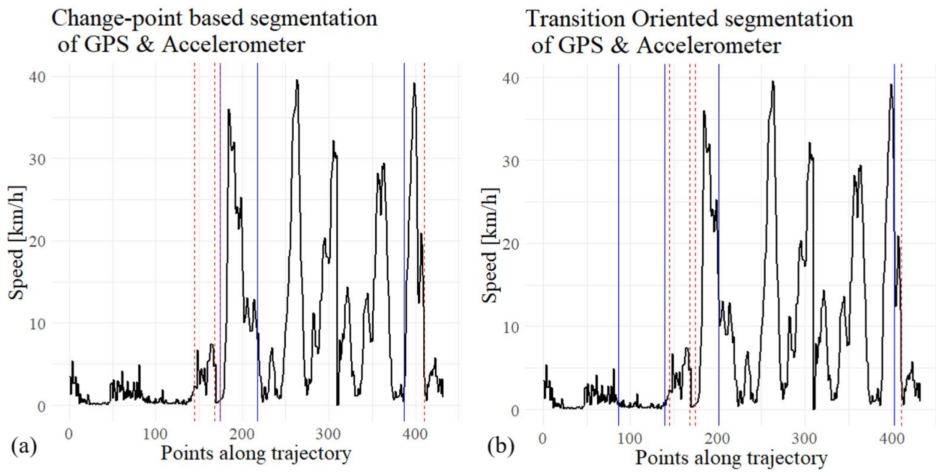

Figure 5.4: Results of the segmentation for a trajectory. (a) shows the calculated transition points for the change-point based segmentation method, and (b) shows the transition points from the transition-oriented segmentation method.

Figure 5.5 shows the problem of over-segmentation of a unimodal trajectory due to the periodic behaviour. For this exemplary trajectory, the recorded transportation mode is tram, described by the variable speed pattern. In using the selected threshold values for the change-point segmentation method (fig. 5.5 (a)), only one transition point is detected. Where the function for the transition-oriented segmentation method detects multiple transition points in the trajectory and heavily over-segments it (fig. 5.5 (b)).

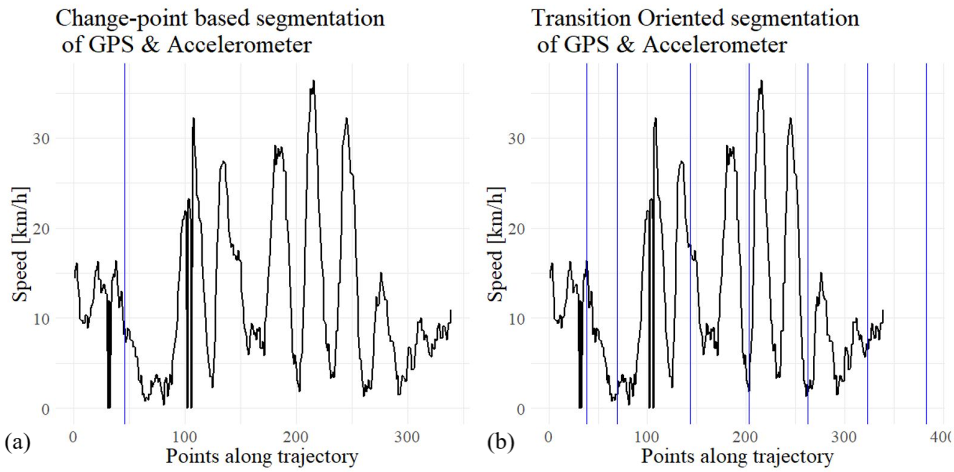

Figure 5.5: Speed profile and detected transition points by both segmentation methods for a unimodal trajectory consisting only of the records of the transportation mode tram. (a) shows the result of the change-point based segmentation and (b) the result of the transition-oriented segmentation.

5.1.1 Influence on classification results

To select the most appropriate segmentation method, the three different segmentation methods discussed in Section 3.2 are applied with classifiers trained by the benchmark dataset. The accuracies are expected to be high, due to the same environment of the training and testing data; nevertheless, a first overview into the usability of the segmentation methods can be gained. The initially applied segmentation methods to train the benchmark data classifiers are segmentation by fixed window size and by labels, which is also tested due to the availability of the ground truth labels. For the validation of the classifiers with the benchmark data, higher accuracy and kappa values are achieved, as can be seen in Table 8.1 (see Appendix 1.1). Following this, segmentation by label is not further discussed in this thesis.

To quantify the results from the segmentation methods, the classification of the segments determined by three segmentation methods is performed, using a fixed window size trained classifier. Table 5.1 shows the overall accuracy of the models with data segmented by the three different methods. Data segmented by a fixed window size yields the highest accuracy values, followed by the change-point based segmentation. For all segmentation methods and sensors, the RF algorithm provides the highest accuracies, using accelerometers' or both sensors' data. The fixed window size classifications are expected to be having the highest accuracy, due to the same classification method for the training and testing the classifiers. But the difference to the other methods is higher than expected, confirming the use of the fixed window size segmentation method.

Table 5.1: Accuracies of classification of benchmark data segmented by the different segmentation methods, using different sensors and different classifiers.

Segmentation method	**Transition-oriented**	**Change-point based**	**FixedTime**
GPS			
Random Forest	57.84%	70.55%	90.19%
Support Vector Machine	51.60%	66.62%	72.42%
Neural Network	47.03%	62.39%	79.78%
Accelerometer			
Random Forest	83.44%	83.35%	93.5%
Support Vector Machine	75.09%	75.03%	92.54%
Neural Network	58.4%	58.26%	82.36%
GPS & accelerometer			
Random Forest	88.22%	89.36%	95.45%
Support Vector Machine	74.71%	82.39%	94.32%
Neural Network	64.37%	68.81%	85.12%

5.1.2 Choice of fixed time window segmentation method

Based on the above-mentioned results of the segmentation, the segmentation method selected for the remainder of the analysis is the fixed window size. This choice is justified for the following reasons:

- In using the other segmentation methods, the accuracies of the classifiers are already lowered based on the choice of segmentation method. As the goal is to study the choice of sensors, classifiers and the application of trained classifiers to new datasets, the use of the change-point based and transition-oriented segmentation method introduces an additional error source.
- Short segments are difficult to detect, as the premise for segmenting the trajectories in the time-oriented and change-point methods was set to a minimal segment length of at least one minute in order to enable a reliable classification. Even in the benchmark dataset, there are quick transitions

between transportation modes, which are shorter than that time frame. In not recognising those transitions, it is not possible to classify additional transportation modes, as entire segments are not recognised. In using the fixed window size, each time interval is labelled by the classifier.

- The dataset onto which the segmentation methods are classified has inadequate temporal coverage. As the sensor coverage is poor, there are many time gaps in the MOASIS data, without even looking at the sensor coverage. This results in difficulties of directly applying the proposed segmentation methods on the data, as they both rely on complete trajectories, such as the ones provided by the benchmark dataset. The algorithms would need to be additionally expanded to include such conditions for the segmentation.

5.2 Feature Selection

The features introduced in Section 4.3 are generated over the defined segments of one-minute length. All selected features were used for the training of the classifiers and their relative importance in a trained classifier was determined using the *varImp()* function from the *caret*-package (Kuhn, 2019). In the following subchapters, the scaled feature importance depending on the classifiers and on the sensors is presented. Table 8.2 (see Appendix 1.2) provides an overview in listing the names seen in the figures and assigns them to the introduced features.

5.2.1 Variable importance different classifiers

In this subchapter, the feature importance is shown for the different classifiers trained by the benchmark dataset, with respect to the features from the different sensors. Figure 5.6 illustrates the variable importance for the RF classifier, which shows the most promising results for transportation mode detection, as seen in Table 5.1. When looking at the variable importance for the GPS sensor (fig. 5.6 (a)), the five most important features are the maximum, 95^{th} percentile, standard deviation, range and the 80^{th} percentile of speed. Features that have low importance in classifying transportation modes are the minimum speed, the sum of the turning angle and the stop time.
For the accelerometer (fig. 5.6 (b)), the five highest values are all derived from the total acceleration, summarising the characteristics over all three axes. Features with the highest importance are the interquartile range, 95^{th} percentile, 80^{th} percentile, 20^{th} percentile and the minimum of the total acceleration. Here two features describe each the highest and lowest values of acceleration characteristics of a segment. The interquartile range describes the acceleration between low and high quartiles. Accelerometer features with the lowest importance are the kurtosis in all three axes and the number of zero crossings in the x-axis.
For many features, the relative importance is lower for the accelerometer signal. For the GPS features, half have an importance over 50, where for the accelerometer features only seven have an importance higher than 50.

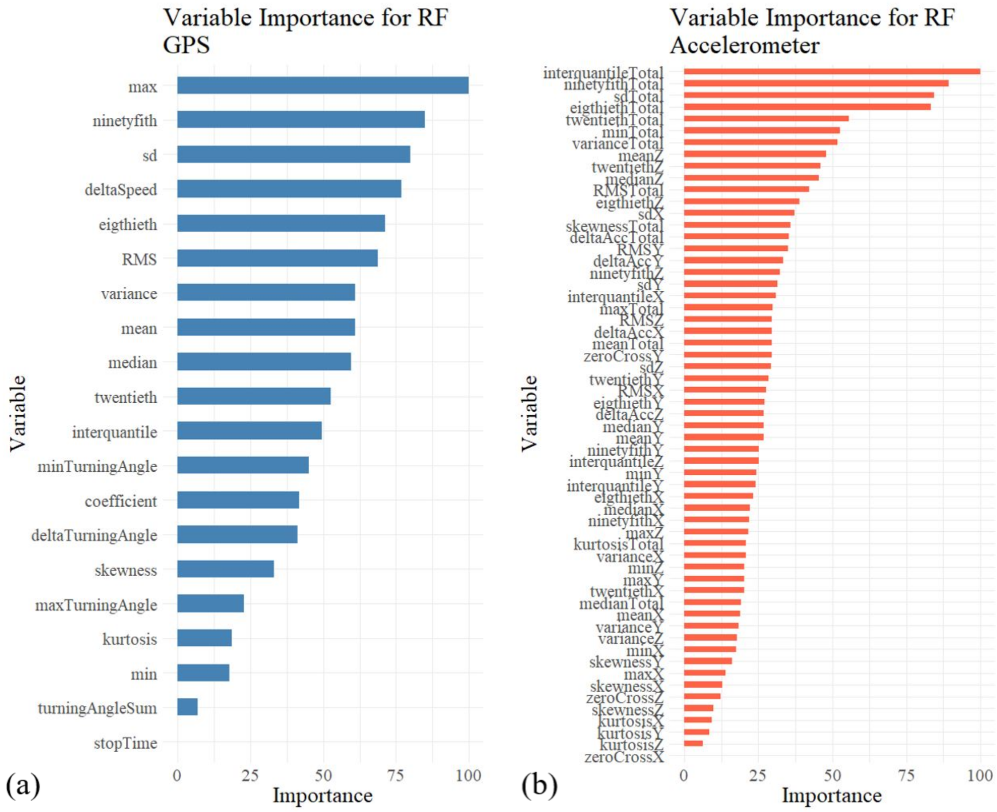

Figure 5.6: Variable importance for RF classifier trained with the benchmark dataset. (a) shows the variable importance of the classifier using only GPS data is and in (b) the variable importance using accelerometer data is shown.

Looking at the combination of features from both sensors in Figure 5.7, out of the ten most important features, nine belong to the GPS sensor. Interestingly, the five most important features from the GPS sensor (95^{th} percentile, maximum, root mean square, range and 80^{th} percentile of speed) have a similar variable importance score as for the classifier trained only with GPS data. Only the RMS gained importance and reduced the importance score of the standard deviation by one position. For the accelerometer features, unlike the results for the single sensor, the highest-scoring variables come from different axes. The three features with the lowest ranking are the sum of turning angles for the GPS and the number of zero crossings in the x- and z-axis.

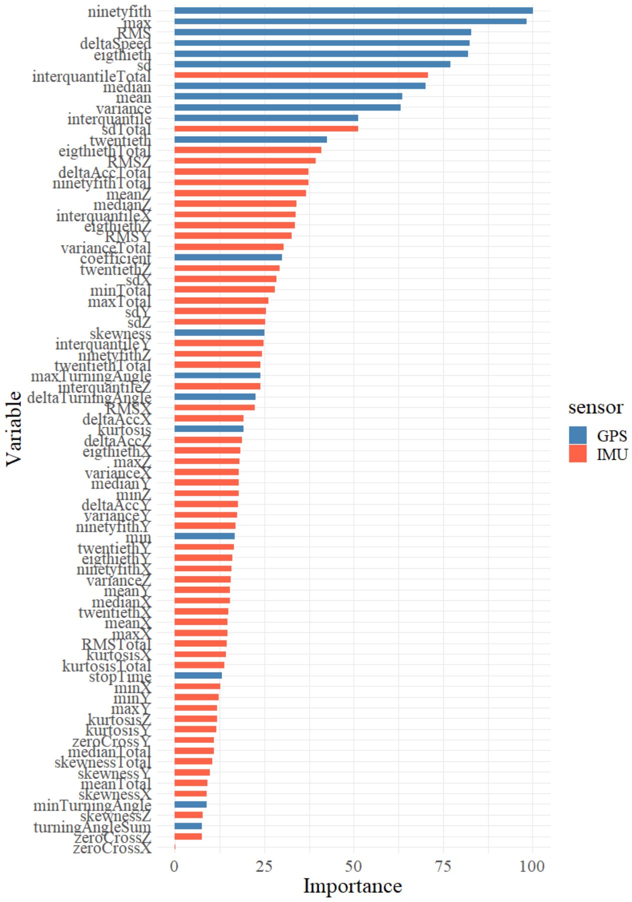

Figure 5.7: Variable importance for RF classifier trained with the benchmark dataset. The variable importance of the classifier using both sensors is illustrated. GPS features are shown in blue and accelerometer features in red.

The variable importance of SVM and the NN classifier, are illustrated in Figures 8.1 and 8.2 (see Appendix 1.3), respectively. Both classifiers trained with GPS data have an identical order in variable importance, only differing in importance scores. For both classifiers, the highest features with the highest importance are the maximum, variance, standard deviation, 95^{th} percentile and range of speed, with the lowest ranking features being minimum, turning angle sum and stop time.
For the classifiers trained using accelerometer data, the highest-ranking features for the SVM are the 80^{th} percentile, interquartile range, 20^{th} percentile, 95^{th} percentile and variance of the total acceleration. For the NN, the highest-ranking variables are the 80^{th} percentile, interquartile range, 95^{th} percentile, variance and standard deviation of the total acceleration. Lowest ranking features for the SVM are mean of the y-axis acceleration, skewness and 20^{th} percentile of acceleration in the x-axis. For the NN, the lowest ranking variables are 20^{th} percentile, median and mean of the acceleration in the x-axis.
When combining the sensors, the ten highest-ranking variables for the SVM are composed of seven accelerometer features and three GPS features. For the NN trained classifier out of the ten highest-ranking features, the composition is the opposite, with seven coming from the GPS and three from the accelerometer features. The five highest-ranking features for the combined SVM classifier are interquartile range, 80^{th} percentile and 95^{th} percentile of total acceleration, interquartile range of the x-axis acceleration and variance of the total acceleration. For the NN the highest-ranking features are the interquartile range,

80^{th} and 20^{th} percentile of the total acceleration and the variance and standard deviation from the GPS speed. The lowest ranking features for the SVM are skewness and median for the acceleration in the y-axis and skewness of the acceleration in the z-axis. For the NN they are median of acceleration in the x-axis and stop time and kurtosis of the GPS speed. Table 5.2 summarises all the features with the highest and lowest importance for all the classifiers and sensors trained with the benchmark dataset to give an overview of all feature importance scores.

Table 5.2: Overview over the features with the highest and lowest importance of all the classifiers trained with the different sensor combinations of the benchmark dataset and by the different classification methods.

Sensor	Random Forest	Support Vector Machine	Neural Network
GPS			
Highest Importance	Maximum, 95^{th} percentile, standard deviation, range, 80^{th} percentile of speed	Maximum, variance, standard deviation, 95^{th} percentile, the range of speed	Maximum, variance, standard deviation, 95^{th} percentile, range of speed
Lowest Importance	Minimum speed, the sum of the turning angle, stop time	Minimum speed, the sum of turning angle and stop time	Minimum speed, sum of turning angle, stop time
Accelerometer			
Highest Importance	Interquartile range, 95^{th} percentile, 80^{th} percentile, 20^{th} percentile, the minimum of the total acceleration	80^{th} percentile, interquartile range, 20^{th} percentile, 95^{th} percentile, variance of total acceleration	80^{th} percentile, interquartile range, 95^{th} percentile, variance, standard deviation for the total acceleration
Lowest Importance	Kurtosis in all three axes, the number of zero crossings in the x-axis	Mean y-axis, skewness z-axis, 20^{th} percentile y-axis	20^{th} percentile, median and mean of the acceleration in the x-axis
GPS & accelerometer			
Highest Importance	95^{th} percentile, maximum, root mean square, range, 80^{th} percentile of speed	Interquartile range, 80^{th} percentile, 95^{th} percentile of total acceleration, interquartile range x-axis, variance total acceleration	Interquartile range, 80^{th} and 20^{th} percentile of the total acceleration, the variance and standard deviation from speed
Lowest Importance	Sum of turning angles for the GPS and the number of zero crossings in the y- and z-axis	Skewness y- and z-axis, median y-axis	Median of acceleration in the x-axis and stop time and kurtosis of the GPS sensor

5.2.2 Variable importance for hierarchical classification

One of the most interesting hierarchical classifications is to distinguish between active and passive modes of transportation correctly. In the following, only results from the RF classifier are shown, due to the highest classification performance obtained from this classifier.

The variables of high importance and their ranking can be seen in Figure 5.8. For the classifier trained by the GPS data (fig. 5.8 (a)), the highest are the 95^{th} percentile, maximum, range, standard deviation and 80^{th} percentile of speed. They are the same variables that were seen in the results for the classification into the seven transportation modes, but in a slightly different order. Variables with low importance are the minimum turning and sum of turning angles in a segment, and the count of points where the individual stopped.

For the classifier trained by the accelerometer (fig. 5.8 (b)), the highest importance for the accelerometer classifier is composed only of variables summarizing the acceleration for the three axes. Those are the standard deviation, 95^{th} percentile, interquartile range, range and 80^{th} percentile of the total acceleration. In comparison to the classifiers for all the transportation modes, the total range of acceleration is of increased importance. Variables with low importance are the mean of the acceleration in the y-axis, the counts of zero crossings of the x-axis and the 95^{th} percentile of acceleration in the y-axis. In comparison to the classifier looking at all transportation modes, the importance of the kurtosis in all axes is increased.

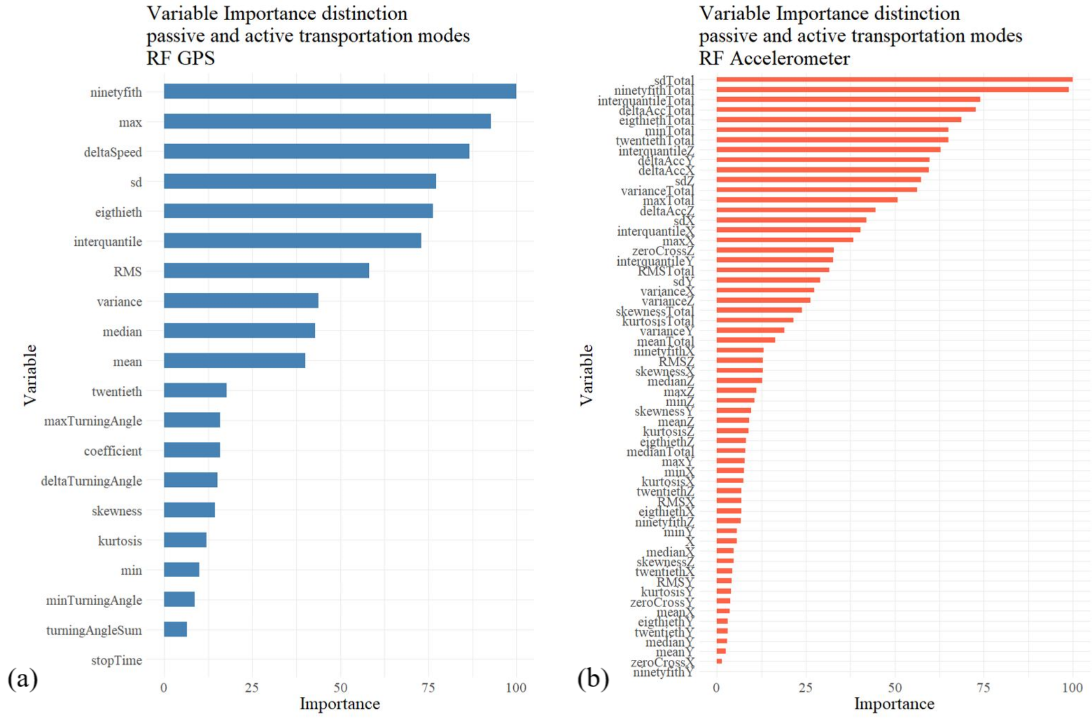

Figure 5.8: Variable importance for RF classifier distinguishing between the two classes of passive and active transportation. (a) shows the results for the GPS sensor classification and (b) for the accelerometer sensor classification.

In Chapter 2, the difficulty of reliably classifying challenging modes of transportation has been introduced. All of them are passive transportation modes, especially the distinction of public transportation remains a challenge. In the following, the variable importance for the classification between different public transportation modes is described and illustrated in Figure 8.3 (see Appendix 1.4), showing the essential features for this challenge. For the classification of public transportation, the variables with the highest importance from the GPS sensor data are the 95th percentile, maximum, 80^{th} percentile, median and RMS of speed. For the accelerometer sensor, the variables with the highest importance are the interquartile range, 80^{th} percentile and 95^{th} percentile of total speed, as well as the median and mean of acceleration in the y-axis. Lowest ranking features for the GPS are the sum, range and maximum of the turning angle. For the accelerometer, the features with the lowest importance are the sum of zero crossings of acceleration in the y- and z-axis, and the kurtosis in all three axes. The variable importance for the classifier using both sensors is shown in Figure 5.9. Here the highest-ranking features are all from the GPS sensor data, and consist of the 95^{th} percentile, maximum, 80^{th} percentile, RMS and median of speed. The highest-ranking accelerometer feature is the interquartile range of the total acceleration. The lowest features are again the turning angle features from the GPS and the count of zero crossings from the accelerometer data.

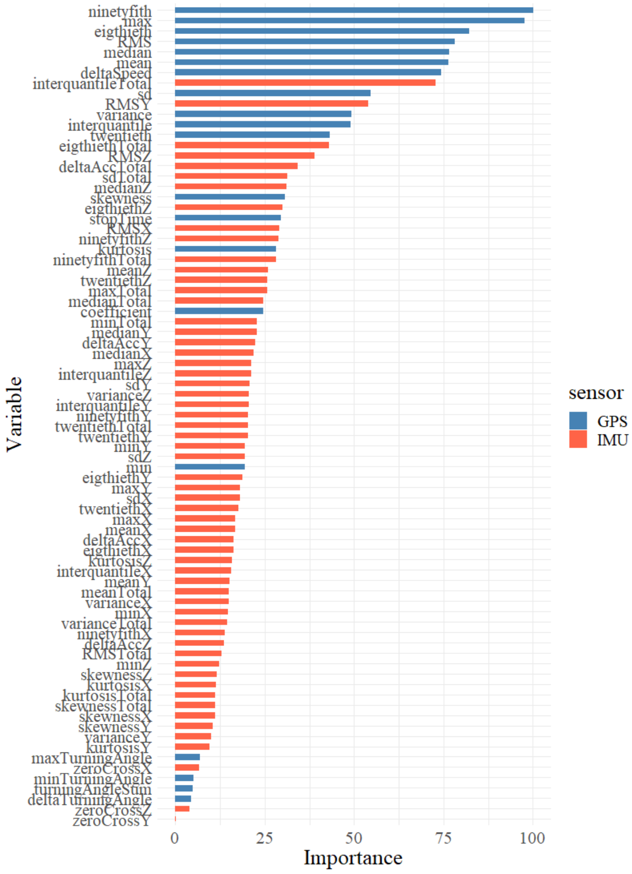

Figure 5.9: Variable importance for the classifier of public transportation using both GPS & accelerometer features. GPS features are illustrated in blue and accelerometer features in red.

5.2.3 Variable importance depending on training data

Towards applying a classifier trained with a different dataset than the one to be tested on, the difference in variable importance is shown in this subchapter. For this, the RF classifier has once each been trained with the benchmark and MOASIS dataset. The order and importance of features depending on the sensors are shown in Figure 8.4 (see Appendix 1.5). The highest-ranking feature for the GPS classified data (fig. 8.4 (a)) is the minimum of speed, the following variables and the lowest ranking variables are similar to the ones from the benchmark dataset. For the accelerometer sensor (fig. 8.4 (b)), the highest and lowest ranking features are very similar to the ones from the benchmark dataset. Figure 5.10 shows the variable importance for the RF classifier trained by the two available datasets using the GPS and accelerometer data. For the data trained with the benchmark (fig. 5.10 (a)), out of the ten highest-ranking features, nine are derived from GPS data and only one from the accelerometer data. Looking at the RF classifier trained with the MOASIS data (fig. 5.10 (b)), all ten highest-ranking features are derived from the accelerometer, with nine describing features from total acceleration and one describing the 80th percentile of acceleration in the z-axis.

For the lowest ranking features, the opposite is the case. Looking at the classifier trained with the benchmark dataset (fig. 5.10 (a)) out of the ten lowest-ranking features, only two are derived from the GPS and eight from the accelerometer. For the classifier trained with the MOASIS data (fig. 5.10 (b)) out of the ten lowest-ranking features, seven belong to the GPS sensor and only three to the accelerometer, which are the count of crossings of the zero-acceleration line in all three axes.

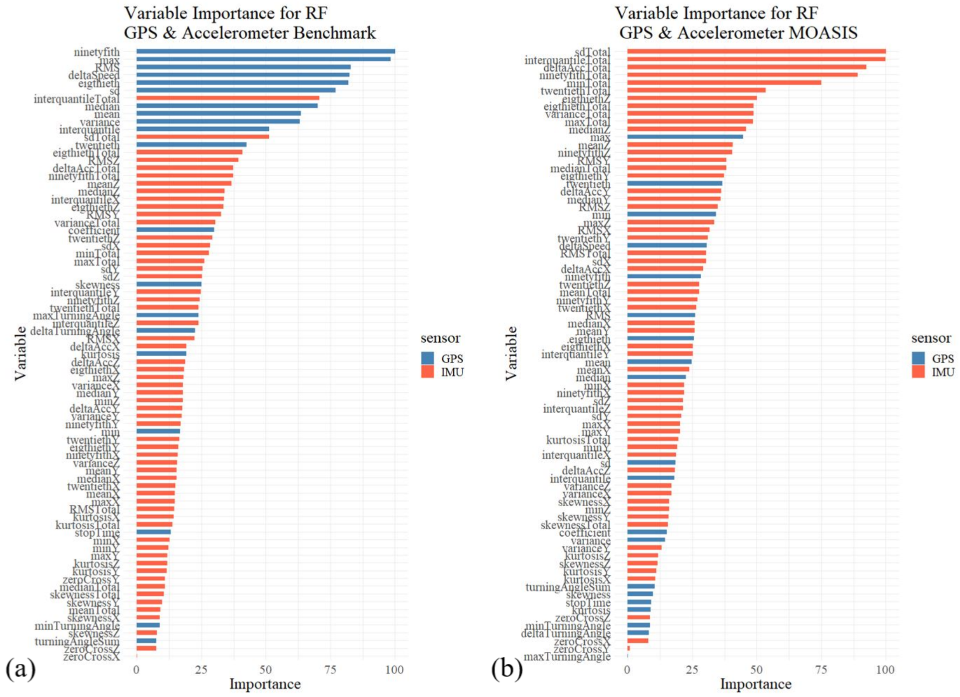

Figure 5.10: Variable importance for the RF classifier trained using both the GPS and Accelerometer features. (a) shows the variable importance for the classifier trained by the benchmark dataset and (b) for the classifier trained by the MOASIS dataset.

5.2.4 Importance of variables with few correlations with velocity

As introduced in Section 4.3, in addition to features describing the character and changes in speed and acceleration, additional features thought to be having few correlations with directly measured velocity are selected for mode detection. In this subchapter, their importance in the trained classifiers is illustrated. The selected features for the GPS data are the stop time and features derived from the change in heading of a segment and for the accelerometer signal, the number of zero-crossings for the three axes are counted. In Figure 5.11, these variables are highlighted for the RF classifier trained with the benchmark dataset. It can be seen that the features not correlated with velocity rank relatively low, with the highest-ranking ones being the maximum and range of the turning angle for a segment. Out of the five lowest-ranking features, four of them were thought to distinguishing challenging transportation modes. Figures 8.5 & 8.6 (see Appendix 1.7 & 1.8), show that irrespective of the sensors and the dataset, the additional selected features are always among the lowest ranking ones, even more so for the real-life MOASIS dataset than for the benchmark dataset.

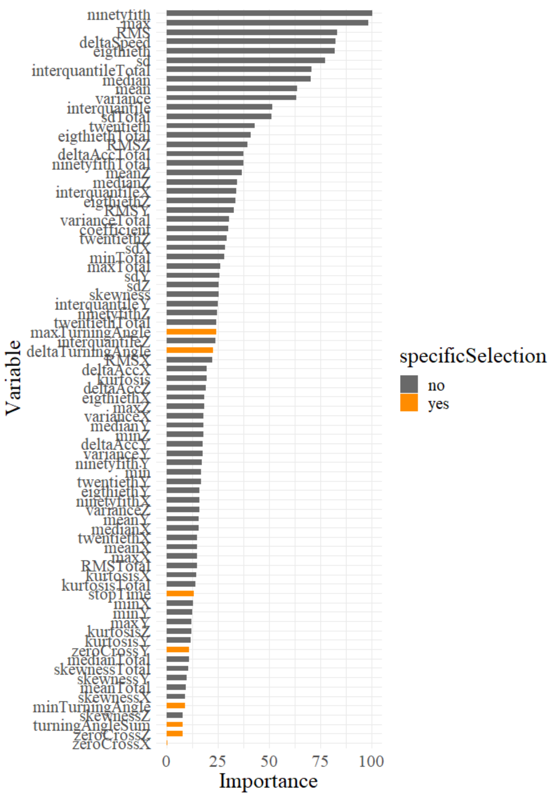

Figure 5.11: Variable importance for the RF classifier trained using the benchmark dataset and both sensors. Specifically selected features for the distinction of difficult transportation modes are highlighted in orange.

As the features are explicitly selected for the distinction of challenging modes of transportation, which consist of motorised and public transportation modes, Figure 5.12 shows the variable importance of those features. Figure 5.12 (a) shows the importance of the RF classifier distinguishing between private and public modes of transportation and Figure 5.12 (b) shows the importance for the classifier distinguishing between the four public transportation modes. For classifying private and public transportation modes, the range of the turning angle is of relatively high importance. The classifier distinguishing between public transportation modes ranks stop time as relatively important. However, for both classifiers, the majority of the specially selected features rank lowest in importance. Having a look at the variable importance for the same classifiers trained with the MOASIS data in Figure 8.7 (see Appendix 1.8) shows that the selected variables all rank lowest, with the exception of the feature counting zero-crossings in the z-axis.

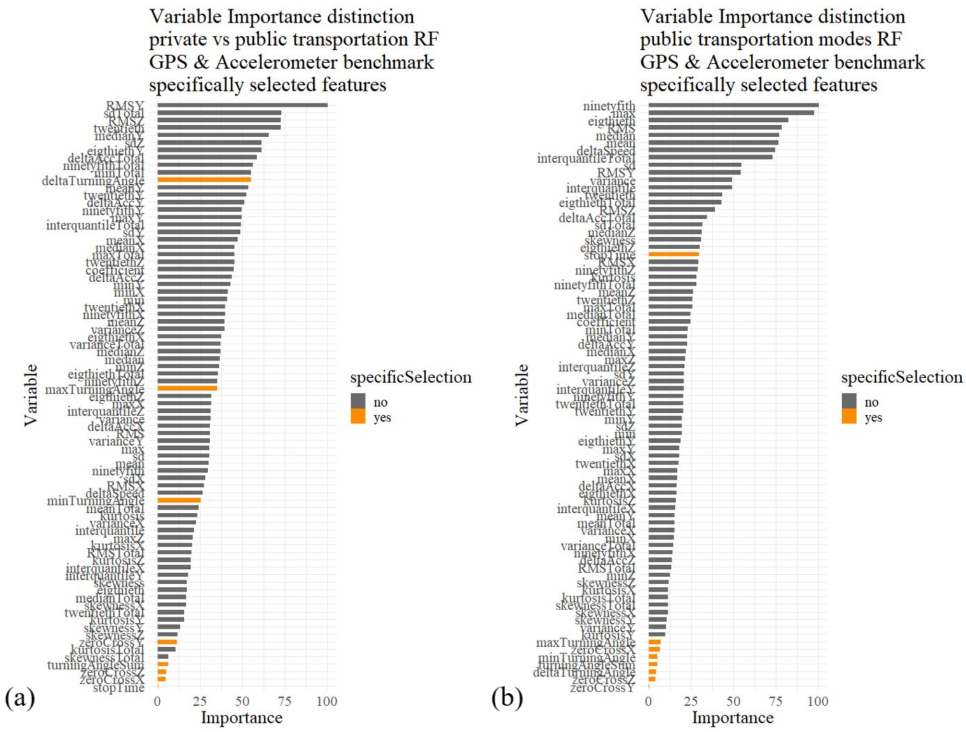

Figure 5.12: Variable importance for the classifiers trained with the benchmark data. The specially selected features are highlighted in orange. On the left (a), the results for the classifier distinguishing between private and public transportation is shown and on the right (b) for the classification of the four public transportation modes.

5.3 Classification

5.3.1 Trained classifiers with benchmark dataset

Table 5.3 gives an overview of the classification results for the three different classifiers trained with the benchmark dataset. The results of the application of a trained classifier to a different type of dataset are not promising. NN produces very low overall accuracy values, but it is the only classifier that is able not to predict only one class for a specific classifier. The RF classifier yields a relatively high accuracy using accelerometer readings, but even though the accuracy value seems high, it is a poor result because all entries are classified as walking segments. The accuracy value results are high only due to the frequent occurrence of walking in the dataset. When using both GPS & accelerometer features, the accuracy value drops slightly for the RF accelerometer, but at least not all entries are classified as the same mode. For the SVM, the accuracies using accelerometer sensor and GPS- & accelerometer sensors for the classification encounter the same problem of all being classified as walking segments. Looking at the kappa values, which describe the classification result better in removing randomness, the highest values follow the best accuracy results but are lowered substantially. Nevertheless, the level of agreement is weak to none existent. Based on the low accuracy values that can be seen in this table, only the RF classifier is taken into consideration for the remainder of the analysis.

Table 5.3: Accuracy and kappa values of the classifications performed with a different training and testing dataset for the three analysed classification methods and the three different sensor combinations. For testing only MOASIS data with a good temporal resolution has been selected.

Classifier	GPS	Accelerometer	GPS & accelerometer
NN			
Accuracy[%]	30.39	13.95	0.0107
Kappa	0.0373	0.0396	-0.0512
RF			
Accuracy[%]	75.8	68.21 (all as walking)	74.54
Kappa	0.4484	0	0.3332
SVM			
Accuracy[%]	60.34	68.21 (all as walking)	68.21 (all as walking)
Kappa	0.2416	0	0

5.3.2 Random Forest classifier trained with MOASIS data

The experiment of training the classifiers with the benchmark dataset and applying them to selected data from the MOASIS dataset does not provide satisfactory results, as illustrated in Table 5.3. Following this, the opposite approach is tried; the classifiers are trained with the data collected within the MOASIS study and tested on the benchmark dataset. The idea behind the creation of the benchmark dataset is a testing dataset that can be used by different researchers (Isler, 2018). In Figure 5.13, the results of the RF trained with the MOASIS data are shown. In blue, the accuracies of using the benchmark data to test the classifier are illustrated and in green the accuracies for using MOASIS data to test the classifier. Lower accuracies can be seen for classifiers using testing and training dataset from different data collections. For the classifier using the GPS sensor, the accuracy for testing with the benchmark dataset is 60.75% and for the MOASIS tested data is 67.57%. These results are in the same magnitude and show a close result for the differently tested classifiers. The accuracies for testing with the benchmark dataset are much lower for the accelerometer and GPS- & accelerometer sensor tested classifiers. The testing accuracy with the benchmark data performs over 70% worse than testing with MOASIS data for accelerometer sensor only. When using both the GPS & accelerometer sensor, the difference in accuracy is around 50%.

Figure 5.13: Accuracies for the RF classifier trained with MOASIS data and different sensors. The classifier is tested with the different datasets, accuracies for the classification of MOASIS data are shown in green and in blue for testing on the benchmark data.

As mentioned in Subchapter 3.5, the kappa statistics takes the random predictors into consideration and gives a more reliable value of prediction and is illustrated in Table 5.4. For the benchmark tested classifier, the kappa value has a level of agreement from weak to none, showing the unstable results of the classifier. When using MOASIS data to test the classifier, the kappa values for the accelerometer and GPS & accelerometer results are moderate to strong, confirming the relatively good classifier. The negative kappa value for the classifier tested with the benchmark dataset using only the accelerometer values proves that this classifier cannot be used.

Table 5.4: Kappa-values for the classifiers trained with the MOASIS dataset and tested on both the MOASIS and the benchmark dataset for the three different sensor combinations.

Sensor	Kappa Benchmark dataset	Kappa MOASIS dataset
GPS	0.4516	0.3388
Accelerometer	-0.2165	0.7562
GPS & accelerometer	0.2106	0.8264

5.3.3 Random forest classification results per transportation mode

Confusion matrices are presented in the following to determine the results of the classifiers for the different modes of transportation. The confusion matrix opposes the predicted and actual classes on different axes, enabling a facilitated overview of classification results per class. For an ideal classifier, the highest values are found in the diagonal. The values are shown as a normalised value of all the segments predicted for one class. This way of normalising the results corresponds to the precision values. As there is a significant difference in data entries recorded per transportation mode, the normalisation over the entire confusion matrix would lead to low values for all transportation modes except for walking, irrespective of the performance of the classifier.

The following confusion matrices display the classifier trained with the benchmark dataset and tested with the MOASIS data for the GPS and GPS & accelerometer. As mentioned, the accelerometer classifier assigns all segments to the transportation mode walking. In Figure 5.14, the confusion matrix for the classifier trained using speed features from the GPS sensor data is shown. The highest correct prediction is achieved for the transportation mode walking, followed by car. The last mentionable correct prediction is for the bicycle class. The highest misclassifications occur for car segments, which are wrongly labelled as tram, train, commuter train, bus and bicycle. Some walking segments are misclassified as tram, bus and bicycle segments.

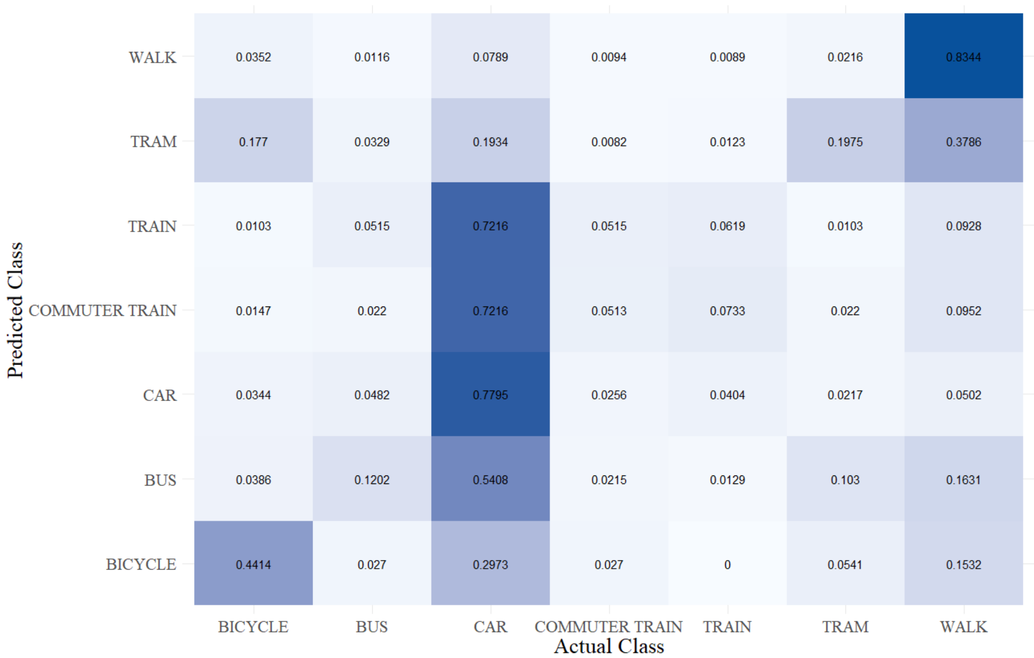

Figure 5.14: Confusion matrix showing precision values for the RF classifier trained with the benchmark dataset using GPS features and applied onto the MOASIS dataset.

In looking at the precision, recall and F1 scores in Table 5.5, it can be seen that when taking the recall into consideration, which includes information about the misclassified data from the transportation mode class, the results are put into perspective. Only the classification of walking segments achieves a very high value of 0.9632. The recall value for the transportation mode car is reduced to 0.423 due to the frequent classification of car segments as the other transportation modes.

Table 5.5: Precision, Recall and F1-score per transportation modes for the RF classifier trained with the benchmark dataset using GPS features and tested on the MOASIS dataset.

Transportation mode (using GPS data)	Precision	Recall	F1
Bicycle	0.4414	0.1231	0.1925
Bus	0.1202	0.1522	0.1343
Car	0.7795	0.423	0.5542
Commuter Train	0.0513	0.1129	0.0705
Train	0.0619	0.0435	0.0511
Tram	0.1975	0.1811	0.1891
Walk	0.8344	0.9632	0.8942

Figure 5.15 shows the confusion matrix for the classifiers trained by the benchmark data and tested with the MOASIS data for the combination of GPS & accelerometer. The highest agreement between prediction and actual class is achieved for the transportation mode car, followed by walking. The highest misclassification occurs for predicting the transportation mode train onto car segments. Generally, as seen in the discussed classifier before using GPS only, the highest misclassifications occur for car segments that are wrongly classified as bicycle, commuter train, train and tram. The agreement for all transportation modes except car and walk is low to non-existent.

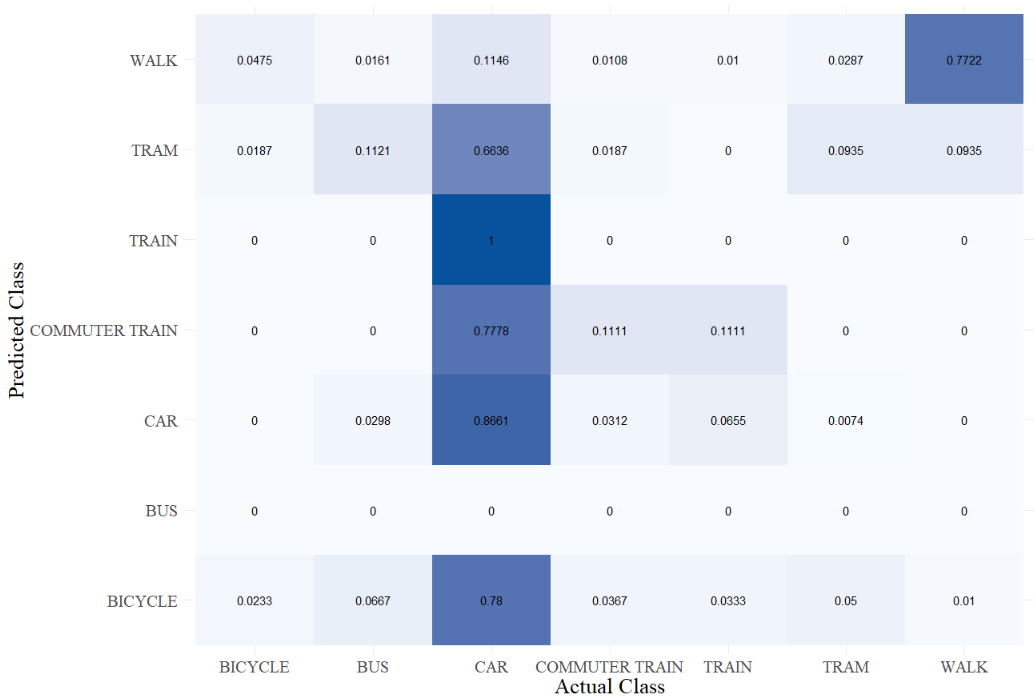

Figure 5.15: Confusion matrix showing precision values for the RF classifier trained with the benchmark dataset using GPS & accelerometer features and applied onto the MOASIS dataset.

Looking at Table 5.6, the high precision values for car segments are opposed by a low recall value. Here the transportation mode with the highest recall value is walking, with 0.9981. Looking at the F1 score, the only transportation mode that is reliably classified is walking; for car, there is a value of 0.4614, and for the other modes, the value is low.

Table 5.6: Precision, Recall and F1-score per transportation modes for the RF classifier trained with the benchmark dataset using GPS & accelerometer features and tested on the MOASIS dataset.

Transportation mode (using GPS & acceleration data)	Precision	Recall	F1
Bicycle	0.0236	0.0176	0.0201
Bus	-	0	-
Car	0.8668	0.3143	0.4614
Commuter Train	0.1176	0.0161	0.0287
Train	0	0	-
Tram	0.0943	0.0377	0.0539
Walk	0.7715	0.9981	0.8703

To put the results of the above-discussed classification into perspective, the results of the classifiers trained and tested using data from the same environment are shown in Figure 5.13. The total accuracy is high for the classifiers trained with accelerometer only (88.71%) and GPS and accelerometer (91.77%). For the GPS only classifier the total accuracy is lower (67.57%).
The confusion matrix in Figure 5.16 reflects the higher accuracies for the GPS trained classifier than for the examples discussed before in this subchapter, with higher values in the diagonal for some transportation modes. The highest agreement between predicted and actual class occurs for the transportation mode walking, followed by car. For the classes bicycle and bus, the correct recognition rate is of around half of the segments, with the highest misclassifications with the actual car and walking segments. Generally, the most incorrect predictions are car and walk segments that are wrongly classified as all the other modes of transportation.

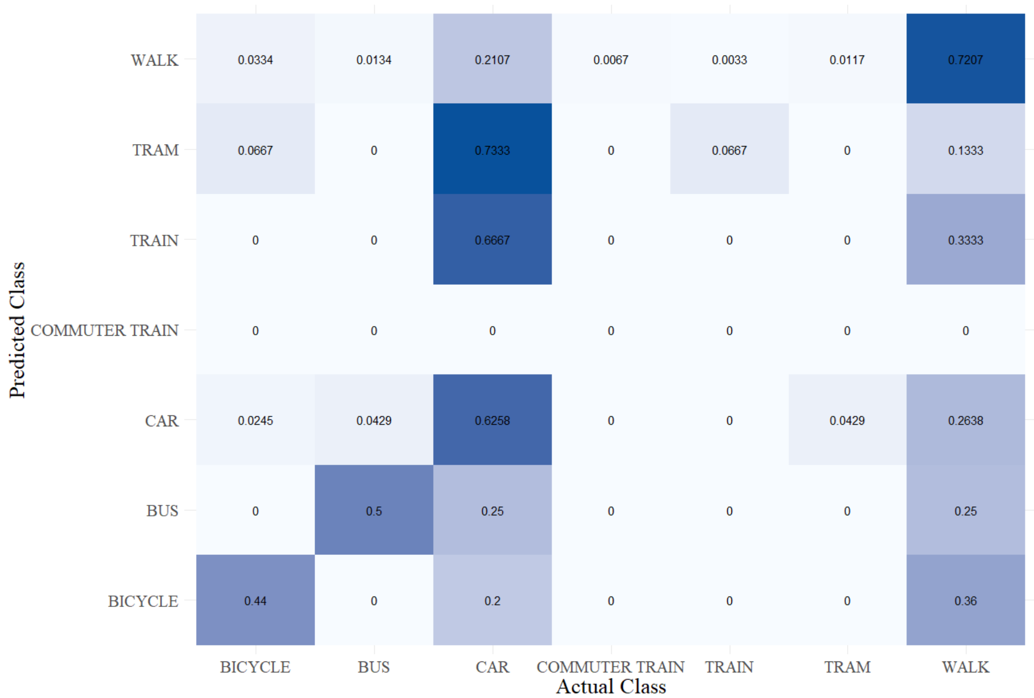

Figure 5.16: Confusion matrix showing precision values for the RF classifier trained and tested on the MOASIS dataset using GPS features.

Looking at precision and recall value Table 5.7, for the classifier trained by GPS only, the best classification result is achieved for the transportation mode walk. Again, the recall of the segments correctly classified into the transportation mode car is slightly lowered due to the misclassification of mostly walking segments as car segments.

Table 5.7: Precision, Recall and F1-score per transportation modes for the RF classifier trained and tested on the MOASIS dataset using GPS features.

Transportation mode (using GPS data)	Precision	Recall	F1
Bicycle	0.44	0.3056	0.3607
Bus	0.5	0.1176	0.1905
Car	0.6258	0.4130	0.4976
Commuter Train	-	0.0	-
Train	0	0.0	-
Tram	0	0.0	-
Walk	0.7207	0.8850	0.7945

The overall accuracy of the classifier trained with accelerometer data is very high; this results in the confusion matrix in Figure 5.17, where the highest values for each predicted class lie in the diagonal, showing a high agreement with the actual class. Proportional misclassification rates are low.

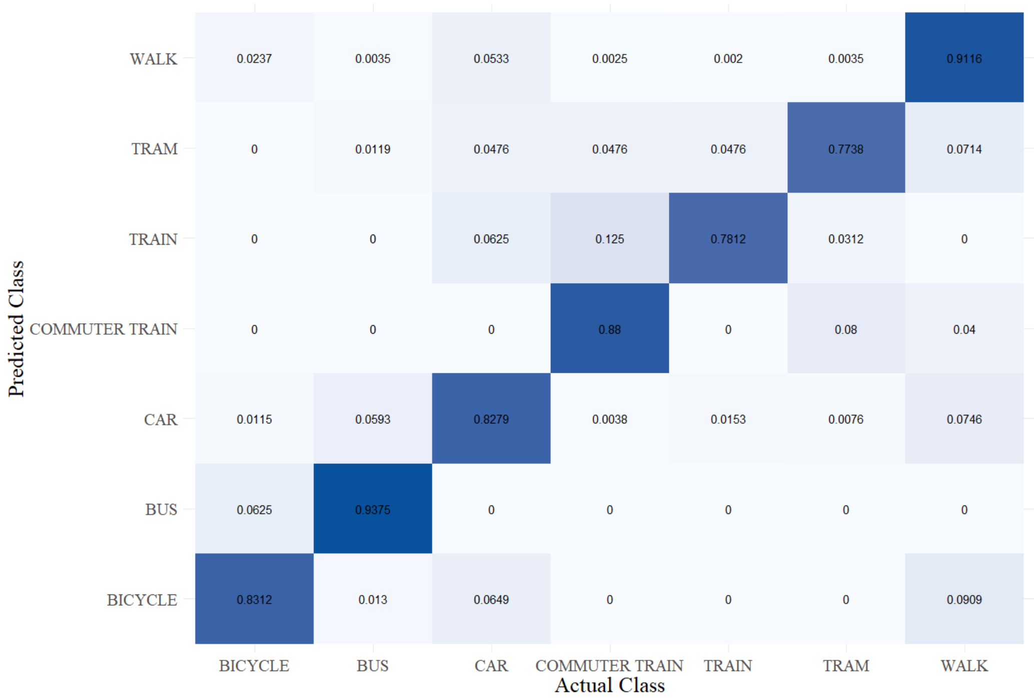

Figure 5.17: Confusion matrix showing precision values for the RF classifier trained and tested on the MOASIS dataset using accelerometer features.

Table 5.8 shows high classification accuracies of using accelerometer data collected in the same study for the training and testing of a classifier in more detail. The precision is over 0.8 for all modes of transportation except for train and tram. These high precisions are put into perspective looking at the recall values, which remain high only for walking, car and tram segments.

Table 5.8: Precision, Recall and F1 score per transportation modes for the RF classifier trained and tested on the MOASIS dataset using Accelerometer features.

Transportation mode (using accelerometer data)	Precision	Recall	F1
Bicycle	0.8289	0.5294	0.6462
Bus	0.9333	0.2545	0.4
Car	0.8279	0.7844	0.8056
Commuter Train	0.88	0.5946	0.7097
Train	0.7813	0.6098	0.6849
Tram	0.7738	0.8228	0.7976
Walk	0.9107	0.9721	0.9404

5.3.4 Hierarchical classification results

For the results of the hierarchical classification, the focus lies on the distinction of passive and active transportation, the distinction into public and private transportation and especially the distinction of public transportation modes. The four different classification steps are introduced in Section 4.4.
First, the results of distinguishing between passive and active modes of transportation are shown in Tables 5.9 & 5.10. In Table 5.9, the four different classification results are shown, listed by the sensor. The classifiers are trained using the benchmark dataset and tested with data from MOASIS. It must be emphasised, that each classifier uses perfect data, meaning that only the transportation modes to be classified are in the testing data. This approach assumes that the previous classification step worked perfectly.
The accuracy and kappa values in Table 5.9, have the highest results for all classifiers when using the GPS sensor. These high accuracies are put into perspective looking at the other kappa-values, where only Classifier 1 achieves a moderate kappa value and the kappa values are minimal for all other classifiers. The classifiers using accelerometer and GPS & accelerometer features have especially low kappa values.

Table 5.9: Accuracy and kappa-values for the hierarchical classification per sensor. The classifiers are trained with benchmark data and tested on the MOASIS data.

Sensor	GPS	Accelerometer	GPS & accelerometer
Classifier 1 - Active vs passive			
Accuracy [%]	85.98	72.5 (all as active segments)	72.73 (almost all as active segments)
Kappa	0.6188	0	0.0126
Classifier 2 - Bicycle vs walk			
Accuracy [%]	94.35	94.09 (all as walking segments)	94.37(most as walking segments)
Kappa	0.3523	0	0.0939
Classifier 3 - Private vs public			
Accuracy [%]	51.12	27.85 (all as public transportation segments)	34.98
Kappa	0.1621	0	0.0421
Classifier 4 - Public transportation modes			
Accuracy [%]	33.61	23.77	28.27
Kappa	0.1324	- 0.0268	0.0408

Looking at the results in Table 5.10, higher values for the classification results are obtained for the classification tested and trained on the same data. The accuracies are over 80% for all the classifiers trained using accelerometer only or GPS & accelerometer features. GPS trained models achieve high accuracies for Classifier 1 and Classifier 2. Looking at the kappa values, the combination of GPS and accelerometer features yields the highest values. Compared to the kappa values displayed in Table 5.9, the results for training and testing classifiers with data from the same collection provides better results.

Table 5.10: Accuracy and kappa-values for the hierarchical classification per sensor. The classifiers are trained and tested on the MOASIS data.

Sensor	GPS	Accelerometer	GPS & Accelerometer
Classifier 1 - Active vs passive			
Accuracy [%]	87.13	92.7	96.15
Kappa	0.6533	0.8131	0.9036
Classifier 2 - Bicycle vs walk			
Accuracy [%]	95.54	96.38	96.98
Kappa	0.4512	0.5797	0.669
Classifier 3 - Private vs public			
Accuracy [%]	77.09	91.1	91.49
Kappa	0.4036	0.7614	0.773
Classifier 4 - Public transportation modes			
Accuracy [%]	58.96	82.55	84.91
Kappa	0.4239	0.7587	0.7895

The following confusion matrices show the results of Classifier 4 in more detail, which distinguishes between the four public transportation modes. As seen in Table 5.9, the accuracy and kappa results are minimal to none for pre-trained classifiers from the benchmark dataset. In the following, the results are analysed in more depth to define the common misclassifications.

For the confusion matrix trained with features from the GPS sensor seen in Figure 5.18, the transportation mode that is most accurately classified is tram. For the three other modes of transportation, always an actual class has been falsely classified to another. Most segments predicted as bus actually belong to tram. For commuter train, there is confusion between train and tram, and for train segments, the highest misclassification happens with bus and commuter train.

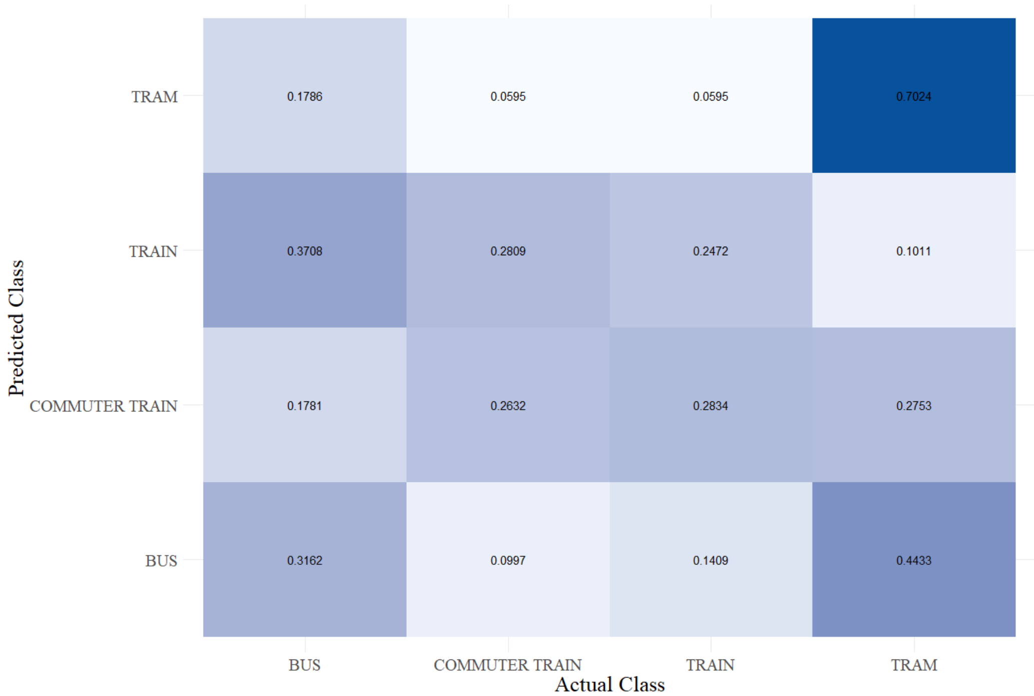

Figure 5.18: Confusion matrix showing precision values for the RF classifier distinguishing between public transportation modes. The classifier is trained with the benchmark dataset and tested on MOASIS data using GPS features.

The relatively high precision value of 0.7 for tram is not reflected in the F1 score displayed in Table 5.11, due to the low recall because many tram segments are misclassified as bus and commuter train. Train has the lowest F1 score, and there is a high misclassification of bus and commuter train segments that are wrongly predicted as train.

Table 5.11: Precision, Recall and F1 score per transportation modes for the RF classifier distinguishing between the four public transportation modes. The classifier is trained with the benchmark dataset and tested on the MOASIS dataset using GPS features.

Transportation mode (using GPS data)	Precision	Recall	F1
Bus	0.3162	0.5	0.3874
Commuter Train	0.2661	0.53226	0.3548
Train	0.25	0.159	0.1947
Tram	0.7024	0.2226	0.3381

The confusion matrix displayed in Figure 5.19, shows the results of the classifier trained using only features from the accelerometer, which has an overall accuracy of 23.77%. No segments are predicted in the commuter train class, and no correct prediction for tram segments resulted from this classification; all segments that are predicted as tram belong to bus or commuter train classes.

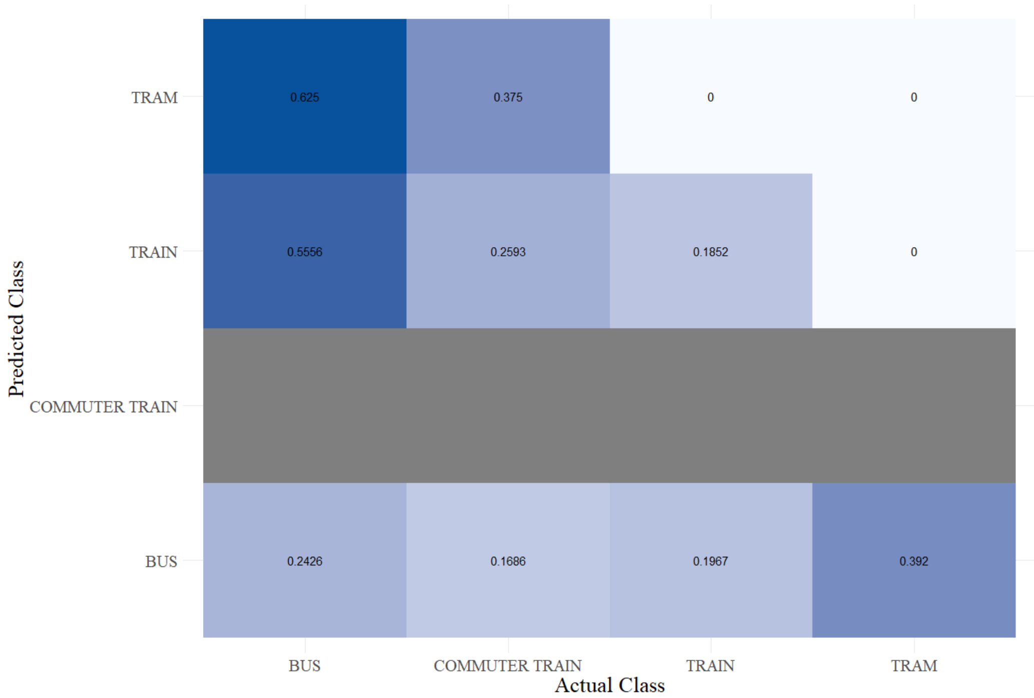

Figure 5.19: Confusion matrix showing precision values for the RF classifier distinguishing between public transportation modes. The classifier is trained with the benchmark dataset and tested on MOASIS data using accelerometer features.

The transportation mode bus has a low precision value, as seen in Table 5.12, which comes from misclassifications, mostly with tram segments. The recall value is relatively high for bus, only a few of the actual bus segments have been misclassified as tram and train segments. The confusion matrix displayed is misleading due to the very different number of transportation modes, and the number of bus segments classified as train and tram is small. In this overall low-performing classification, the transportation mode that is best classified is bus.

Table 5.12: Precision, Recall and F1 score per transportation modes for the RF classifier distinguishing between the four public transportation modes. The classifier is trained with the benchmark dataset and tested on the MOASIS dataset using accelerometer features.

Transportation mode (using accelerometer data)	Precision	Recall	F1
Bus	0.2426	0.8913	0.3814
Commuter train	-	-	-
Train	0.1852	0.0362	0.0606
Tram	0	0	-

The following confusion matrix displays the classifier with the best results, which is trained and tested using MOASIS data with features computed from the GPS & accelerometer sensors. The results of the RF classifier have been selected to display, in order to see where the highest misclassifications remain. With the highest values in the diagonal, the high accuracy of the classifier is seen in Figure 5.20. The correct agreement between predicted and actual classes is the case for all four transportation modes. The highest remaining misclassification is for predicted train segments that are commuter train segments and train segments that are wrongly predicted as bus.

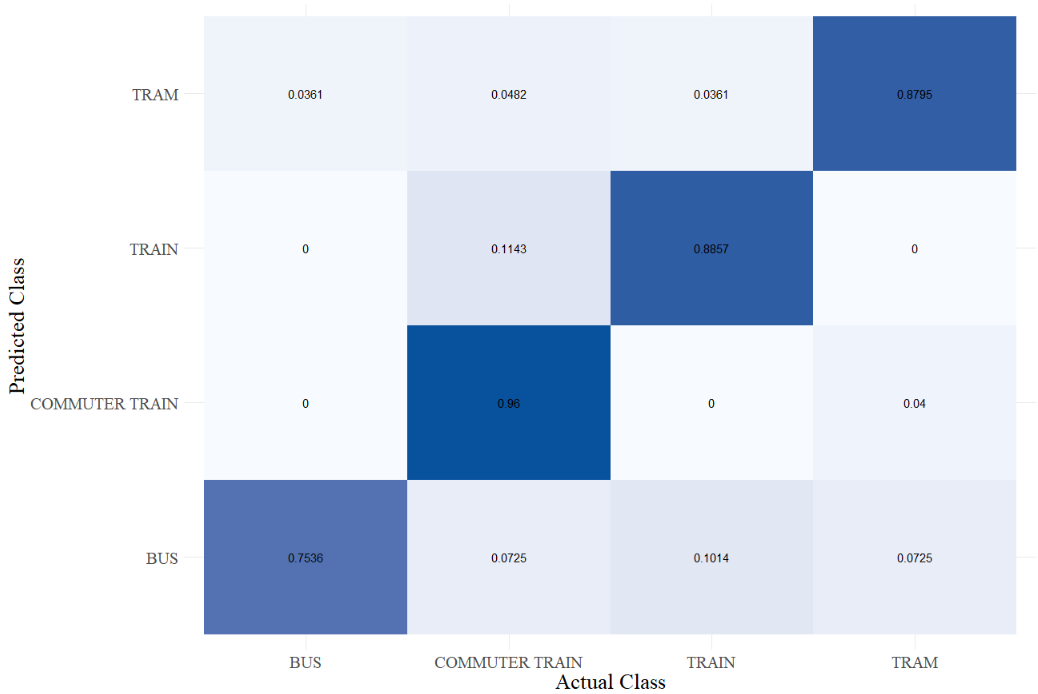

Figure 5.20: Confusion matrix showing precision values for the RF classifier distinguishing between public transportation modes. The classifier is and tested with the MOASIS data using GPS & accelerometer features.

Looking at Table 5.13, the recall values are high for bus and tram, meaning that not many segments of the actual class were classified as another mode of transportation. For commuter train and train there is a lower recall value, commuter train segments are mostly misclassified as train segments, and train segments are mostly misclassified as bus segments. For both, the proportion of misclassification is low. Tram has the highest F1 score, followed by bus, train and commuter train.

Table 5.13: Precision, Recall and F1 score per transportation modes for the RF classifier distinguishing between the four public transportation modes. The classifier is trained and tested on the MOASIS dataset using GPS & accelerometer features.

Transportation mode (using GPS & accelerometer data)	Precision	Recall	F1
Bus	0.7536	0.9455	0.8387
Commuter Train	0.96	0.6486	0.7742
Train	0.8857	0.7561	0.8158
Tram	0.8795	0.9241	0.9012

5.4 Sensor influence

The focus of this thesis is not only to test the transferability of classifiers on a new dataset but also to highlight the difference of results depending on which sensor data is used for building the classification model. In the following, the performance difference depending on the sensors use is illustrated by different evaluation indicators.

Using the accuracy metric, a first overview of the performance difference of the classifiers based on the chosen sensor is pictured. In Figure 5.21, accuracy is lowest for the NN classifier. Disregarding the low values, the highest accuracy of 30.39% comes from using features from the GPS sensor data for classification. For the RF classifier, the accuracies are much higher; here, the highest accuracy comes also from using the GPS sensor data, followed by the combination of both sensors and the accelerometer data. Looking at the SVM classifier, the accuracies for classifiers using the accelerometer only and the GPS & accelerometer features are the same, with lower accuracy for using the GPS sensor data.

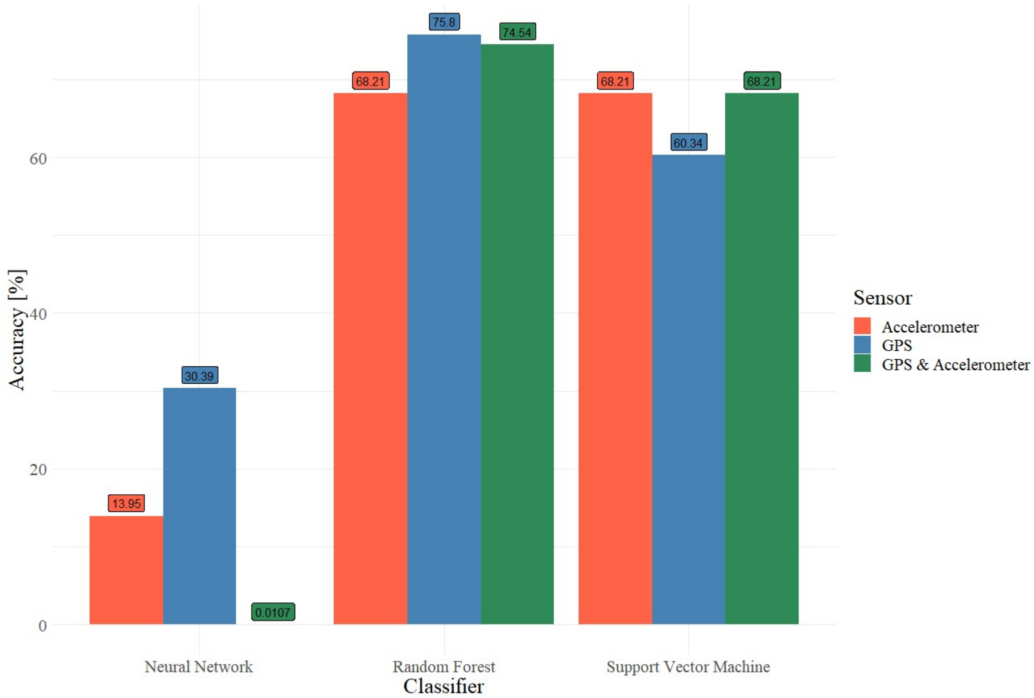

Figure 5.21: Accuracy values for the three tested classifiers per sensor based on training on the benchmark dataset and testing on the MOASIS dataset.

To put the accuracies into perspective, Table 5.3 lists the kappa values for the above-illustrated accuracies. Where the kappa value is zero, all the segments are classified as walking segments. Due to the imbalanced data and the nature of mobility, there are many recorded walking segments, compared to other transportation modes. This results in high but meaningless accuracies. In comparing the kappa values, they are highest for classifiers trained with GPS features. To compare the influence of the selected sensors and their performance per transportation, the F1 score is used to illustrate the results quantitatively.
Figure 5.22 illustrates the results for the RF classifier trained using the MOASIS data and tested using the benchmark dataset. There are no segments classified as commuter train and train, irrespective of the sensor. For bus, tram and bicycle there are only predictions when using the GPS sensor data. Using the accelerometer-based classifiers, all segments are classified either as the transportation modes car or walk. Looking at car and walk – the only transportation modes classified by all three sensors – the performance of using accelerometer features is lowest. This includes the misclassifications, as all transportation modes are classified into two out of the seven classes. The F1 score of GPS features is higher for car than for the combined GPS & accelerometer classification, and for walking segments, it is the

opposite.

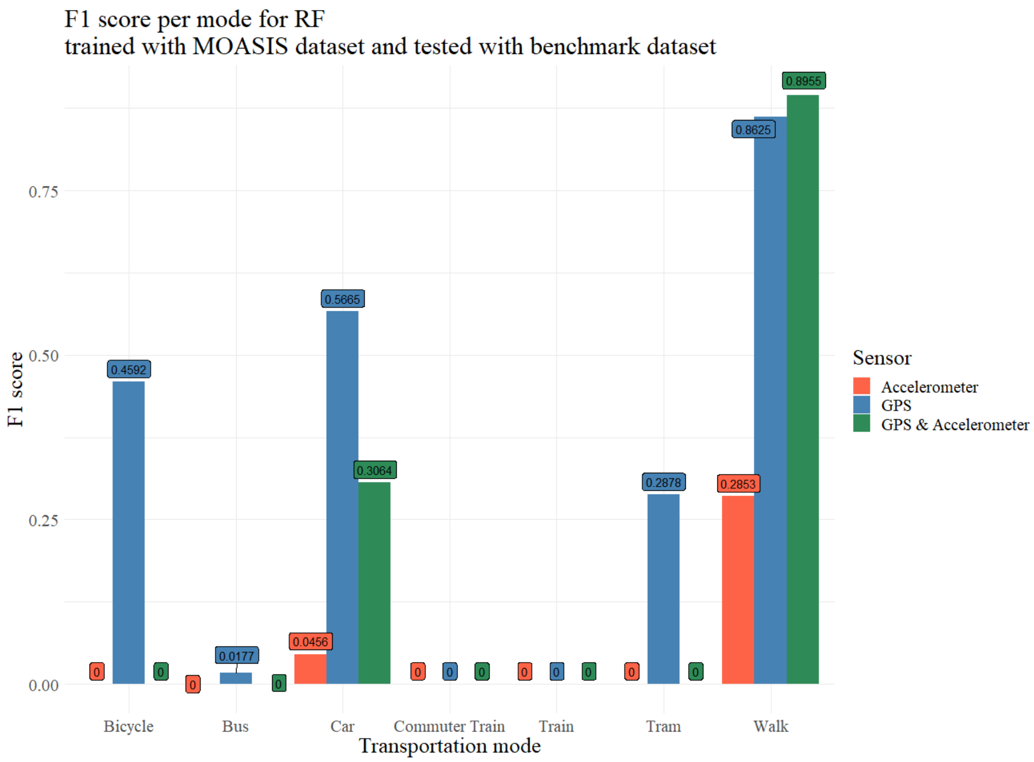

Figure 5.22: Illustration of the F1 scores per transportation mode and sensor for a classifier trained with the MOASIS dataset and tested on the benchmark dataset.

As the transferred classifiers' results are not promising, different F1 scores depending on the transportation mode and sensor data are illustrated for the best classifiers in Figure 5.23. The results come from the RF classifier trained and tested using data from MOASIS. No segments are classified as commuter train, train and tram when using GPS features only. Looking at the transportation modes where there are predictions by the classifiers using features from both sensors, GPS-based models have a lower F1 score than the other two models. For the classifiers trained with features from the accelerometer and both the GPS & accelerometer, the results are similar, with a slightly higher F1 score when using a combination of both sensors data. Commuter train is the only transportation mode where this statement is not valid. The closest F1 score between accelerometer- and GPS & accelerometer-trained classifiers is for the transportation mode bus and bicycle, and the highest difference in the F1 score between these two sensor combination models is for car.

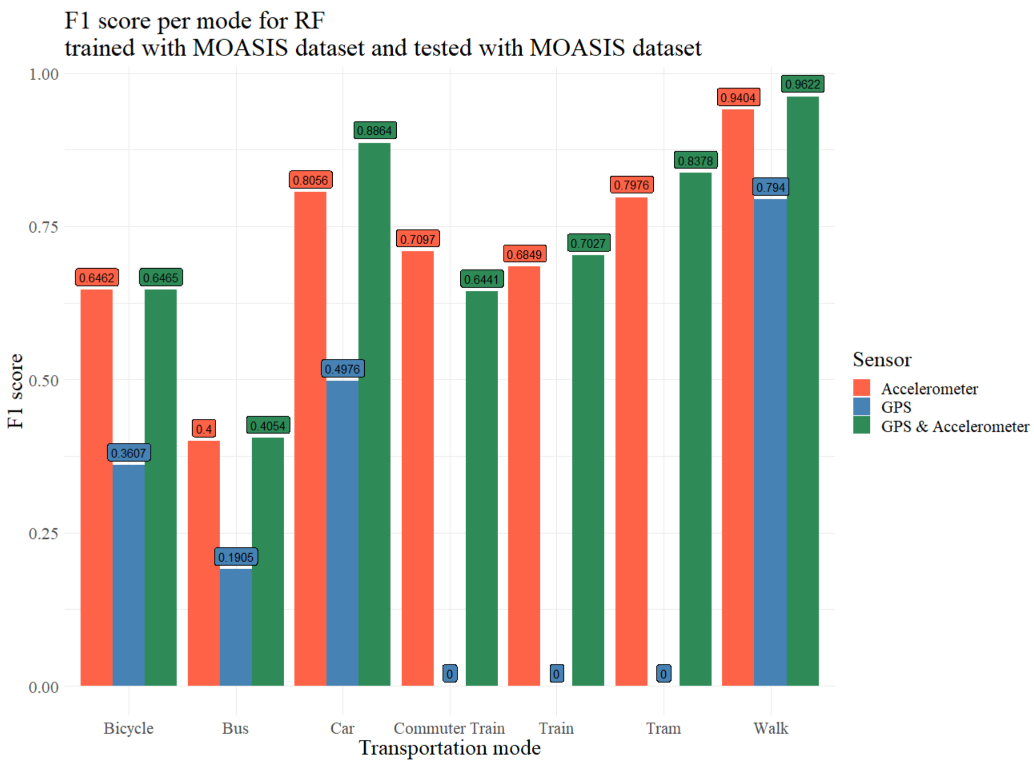

Figure 5.23: Illustration of the F1 scores per transportation mode and sensor for a classifier trained and tested with the MOASIS dataset.

To get an overview of the general performance of the classifiers, the accuracy and kappa values are listed in Table 5.14. The accuracy is lowest for the GPS sensor and increases by over 20% when using a classifier with accelerometer features. The increase in accuracy from using accelerometer only- to GPS & accelerometer-trained classifiers is lower. Looking at the kappa value, for GPS, it is only minimal with a value of 0.3388 and thus substantially lower than the accuracy. For the accelerometer and GPS & accelerometer classifiers, the kappa value is also lower than the accuracy value but has a moderate to strong level of agreement, showing that these classifiers perform better.

Table 5.14: Accuracy and kappa values for the RF classifier trained and tested with the MOASIS dataset.

Sensor	Accuracy [%]	Kappa
GPS	67.57	0.3388
Accelerometer	88.71	0.7562
GPS & Accelerometer	91.77	0.8264

6 Discussion

In this chapter, the results presented in Chapter 5 are summarised and used to answer the two research questions. At the end of the research question subchapters, a summary with the most important information can be found. At the end of the chapter, a critical evaluation is about the performed analysis is made, discussing weaknesses and improvements.

6.1 Research Question 1 – difference in classifier performance depending on sensors

6.1.1 Segmentation

For trajectories where the transitions and characteristics between transportation modes are distinct, both the change-point based and transition-oriented segmentation method can reliably segment the trajectory (fig. 5.1 (b) & (c), fig. 5.2 (b) & (c) and fig. 5.3 (a) & (b)).
The change-point based segmentation selects transitions in the trajectory where selected features fluctuate over a certain threshold, in order to distinguish the change between walking and other modes of transportation. Where the transitions are clear, the changes are well detected. The disadvantage for this segmentation method using simple rules is that all transitions over a certain threshold are recognised, disregarding the regular occurrence of signals, such as the consistent stops of a trip with public transportation. A new segment is always started every time a big enough transition occurs in the characteristic of two consecutive points.
The transition-oriented segmentation method uses an existing algorithm partitioning the trajectory based on similarity within a segment. It is more sensitive to smaller fluctuations and distinguishes between different walking styles within a segment, due to the detection of homogenous behaviour on a smaller scale.

Regarding the use of different sensors for segmentation, it was observed that the best results are achieved when using information from both the GPS & accelerometer sensors. When using data from the GPS sensor, the change-point based segmentation method over-segments the trajectory, while detecting the real transition points well. The transition-oriented segmentation method performs slightly better, showing less over-segmentation. When using data from the accelerometer sensor, the change-point based method under-segments and misses transition points. The transition-oriented segmentation method performs better in detecting homogenous segments that are not separated by large fluctuations. The difference comes from the features used to segment the trajectories. Change-point based segmentation is based on standard deviation and the difference between the maximum and minimum value over a time window of one minute, whereas the transition-oriented segmentation method makes use of the mean and variance of consecutive points. This difference makes the change-point based segmentation slightly more sensitive to fluctuations in the signal.

From the analysis of the data, an option for transportation mode detection is to focus on the middle part of the segments and disregard the features collected during the transition of transportation modes. In visualising the results of the segmentation, it was found that the recorded transition points cannot be recognised by algorithms, due to no signal change. For example, when an individual enters a public transportation mode which stands still for a period before starting, there is no change to be detected by an algorithm. Thus, the transition points calculated by the algorithms, detect recordable changes in behaviour. This statement leads to the focus not being on finding the transition points as close as possible to the recorded ones, but the transition points based on the characteristic features of a transportation mode.

In the literature, few results on trajectory segmentation are presented, with validation done in directly classifying the segmented trajectories. Towards the characteristics of the segmentation process, the five defined features in this thesis are not comparable to the proposed approach by Guo et al. (2012), where 70 features are computed from the accelerometer frame. The most meaningful features are then selected using a principal component analysis. In using an adaptive feature selection model that selects the features and models in optimizing the fit of the data, they achieved a good performance.
Looking at the classification performance following the different segmentation methods in this thesis, the highest is achieved by the fixed window size segmentation method (tab. 5.1). Following the reasons for the fixed window size segmentation to be the most suitable method discussed in Section 5.1.2, its biggest strength is the reduction of error propagation. For the other segmentation methods, the missing of a transition point changes the characteristics of a segment strongly, leading to a signal influenced by different transportation modes which is difficult to classify. With this method, the assumption made for segmentation is removed as it does not necessarily describe real-life behaviour (Ellis et al., 2014). In addition, for the segmentation methods, it must be highlighted that the goal is not to get as close as possible to the labelled changes, but to the recordable changes. This conflict is also removed when using a fixed window size segmentation method.
For all three segmentation methods, the accuracies are lowest when using only the GPS sensor, and increase in using the accelerometer and GPS & accelerometer for segmentation and classification. The highest accuracies are obtained using the RF algorithm, where the difference in accuracy decreases when using the accelerometer or GPS & accelerometer for the segmentation and classification. It must be kept in mind that there is a bias in the classification as the training and testing data come from the same dataset, but the goal in this part is to find out how well the segmentation methods perform. Zheng et al. (2010) apply the change-point based segmentation method using a Decision Tree with GPS data and achieve an overall accuracy of 71%. The results for the change-point based segmentation and RF classification approach applied in this thesis, achieve an accuracy of 70.55% when using only GPS, increasing the accuracy to 83.35% when using the accelerometer. When both sensors are used, the overall classification accuracy is 89.36% (tab. 5.1). These accuracy values show the potential for this method that is applied using simple statistical measures, performing better than the transition-oriented segmentation.

6.1.2 Feature selection

The feature importance for the different classifiers trained by the benchmark dataset shows that for the GPS sensor, the highest-ranking features describe the upper values and range of speed recorded in a segment, characterizing the movement behaviour by the peak and minimum values. For the accelerometer all classifiers use information about the total acceleration, summarising the signal of the different axes and removing the influence of the sensor's orientation. When using features from both the GPS & accelerometer sensor, the RF classifier ranks GPS features highest, using features of the upper percentiles to classify transportation modes. For the SVM only acceleration features are used, describing the maximum and range of total acceleration. SVM is the only classifier that ranks the acceleration of one axis in the five highest features. The NN uses features from both sensors, which describe the range and the highest, as well as the lowest feature values (tab. 5.2).

As seen in Section 5.3, the results from the classifiers trained with the benchmark dataset do not perform well on the MOASIS data. Thus, in the following, some essential variables for the RF classifier trained with MOASIS data are listed. For the GPS the same features are of high importance as from the benchmark trained classifiers, except that the 20^{th} percentile of speed replaces the standard deviation. Looking at the importance scores, for GPS features of the MOASIS trained classifier, the importance of lower ranking-features is higher than for the benchmark trained data. This behaviour reflects the real-life dataset where the distinctions between the different modes of transportations may not be as ideal as for the benchmark dataset. For the accelerometer classifier, the standard deviation and range of acceleration gain importance and the 20^{th} and 80^{th} percentile of total acceleration lose importance. When trained with GPS & accelerometer, the most important features for the RF classifier are all coming from the accelerometer. The classifier trained and tested with MOASIS data does have higher kappa values, but the variable importance is similar to the benchmark-trained dataset except for the case when the sensors are used jointly.

Huss et al. (2014) state that the highest transportation mode classification can be achieved by using GPS features including the 95^{th} percentile of speed, and the difference between the maximum and minimum speed showing the acceleration and deceleration in a segment. For the GPS trained classifiers in this thesis, the usefulness of these variables is confirmed, with them ranking high, irrelevant of the classification performance. It confirms that these features can best discriminatively distinguish between the transportation modes. The range of acceleration suggested by Shafique & Hato (2014) is not found within the high scoring features, but the interquartile range, which is a description of the range of values without outliers, scores high for all classifiers. Ellis et al. (2014) find that the greatest change in error is detected when standard deviation and average speed are used to create new branches in the Decision Tree. For the classifiers trained with the benchmark data, the standard deviation is an essential GPS feature. Martin et al. (2017) find that when using both GPS & accelerometer features, the speed features from the GPS are more useful for the RF classifier. When the benchmark dataset is used, this statement is confirmed; for the use of the MOASIS classifier, the opposite is the case and only features from the accelerometer rank highest.

The features with the lowest score are similar for all the classifiers irrelevant of the sensor. For the accelerometer features, the lowest ranking features are all describing a single axis, highlighting the added importance of the total acceleration value. The skewness and kurtosis mentioned in multiple previous studies as summarised by Pires et al. (2017), and applied for transportation mode detection using GPS features by Xiao et al. (2017), with the skewness ranking in the middle and the kurtosis not on the feature importance scale. Similar to literature, in this thesis, these features are not relevant for transportation mode detection. Other GPS features ranking low in the trained classifiers are various features summarising information about turning angles, computed from the heading change rate in a segment and the feature describing the stop rate within a segment. The feature turn angle is ranked high in importance in literature; the stop rate is found in the middle (Xiao et al., 2017). The results of the trained classifiers do not confirm the usability of these features. Similarly, the count of the zero-crossing rate for the accelerometer sensor in the three axes, that is mentioned in six studies reviewed by Pires et al. (2017) is not found useful for classification.

6.1.3 Classification

Looking at the results from the classifier trained using the benchmark dataset and applied for classification of MOASIS data, the accuracies for the three tested sensor combinations are all highest for the RF classifier (tab. 5.3). As described in Chapter 2, the RF classifier is often used for transportation mode classification and provides reliable results. Even though the size of the training data is modest, the property of the RF to build different classifiers in including different features of the data and deciding the final prediction with a plurality vote is a significant advantage of this classifier towards transportation mode detection (Martin et al., 2017).
In comparison, the characteristics of the SVM classifier are the determining of the best decision boundaries between different classes by a hyperplane that has the most significant margin to points in the classes (Lin et al., 2002). In the classifications performed in this thesis, there are many similar features for the different transportation modes, and it is difficult for the classifier to determine adequate hyperplanes. This results in a lower accuracy and kappa value for the GPS classifier. The classification with accelerometer features result in the prediction of all segments as walk and is therefore unusable. This shows the limitations of this classifier, to achieve better results all transportation mode classes should have the same amount of data to train and test (Shafique & Hato, 2014).
The NN performs best on the GPS trained classifier, continuously reducing the performance for the accelerometer and the GPS & accelerometer classifiers. NN can learn subtle differences from large amounts of data and build reliable classifiers. The poor performance is explainable by the small amount of available data for training, as generally better results are obtained when the monitoring duration is longer and a high sampling frequency is available (Byon & Liang, 2014).

For all benchmark trained and MOASIS tested classifiers (tab. 5.3), the highest accuracy and kappa values are obtained when using GPS features, with an accuracy of 75.8% and a kappa value of 0.4484. Walking segments are being correctly classified the most frequently, followed by car segments. For using GPS & accelerometer features combined, the accuracy is very close to the classifier using GPS features alone, but the kappa value is lowered by 0.11. Classification using the accelerometer seems to be more affected by the data imbalance. For this classification, the statement by (Feng & Timmermans, 2013) that accelerometer sensor classification outperforms GPS sensor classification is rejected. Walking segments are again the highest class that is recognised correctly, followed by car. When using the accelerometer, the accuracy is 68.21%, but as all segments are classified into the walking class this proves why looking at the overall accuracy alone does not describe the quality of the results. The kappa values correspond to zero, as the random predictor is taken into consideration, which influences the accuracy of these cases.
As the results from the classifiers trained with the benchmark dataset and applied on MOASIS data are not sufficiently reliable, the inverse process is implemented. This results in an overall picture the same as the previous classification. In looking at both the accuracy and kappa values (fig. 5.13 and tab. 5.4), the classifier with the GPS features shows the best results for transportation mode detection, followed by the classifier trained with GPS & accelerometer features. The accuracy and kappa values for the classifier trained with accelerometer features are low, as all segments are classified either as the transportation mode car or walk. In prior literature, the use of the accelerometer is expected to improve transportation mode detection (Bedogni et al., 2012). For the classifications that are conducted using classifiers trained and tested by data from different datasets, the opposite is observed within this thesis. The classification using only accelerometer features heavily misclassifies transportation modes into the two classes occurring most often in the dataset. When using both GPS & accelerometer, the influence of the accelerometer reduces the higher accuracy and kappa values from the classifier trained

with GPS features. The more fine-grained information provided by the accelerometer should help to distinguish transportation modes on a smaller scale (Hemminki et al., 2013). This statement cannot be confirmed for the case where the training and testing dataset are collected independently, due to the differing characteristics of the datasets. For the training and testing dataset from the same environment, the accelerometer increases the accuracy of the classification. It appears that the accelerometer sensor is more sensitive towards the characteristics of transportation modes following different data collections.

Looking at the results of the RF classifier trained and tested with the MOASIS dataset (fig. 5.13 and tab. 5.4), the results are more similar to those found in previous literature. The accuracy and kappa values are lowest for the classifier using GPS features and substantially increase when using accelerometer features. The combined classifier with GPS & accelerometer features provides the highest accuracy and kappa value. Looking at the overall accuracy, the increase from using GPS features to accelerometer features is 20%, with a further increase of 3% when using both sensors data combined. This increase is confirmed by the statement of Xia et al. (2014), that accuracy increases by 10% to 20% when using multiple sensors.
In the following, the results are put into relation with accuracy values from transportation mode detection studies using the RF classifier. Due to the composition of the studies found in the literature that use data from the same environment for training and testing, the values are compared to the ones of the classifier trained and tested with MOASIS data. The classification using GPS features only has an accuracy of 67.57% (tab. 5.14). Looking at the precision accuracy, which is the mean of the precision values for all modes of transportation the validity of results drops to 32.65% (tab. 5.7), which is much lower than the precision accuracy of below 76% reported by Stenneth et al. (2011). The low precision for the classifier resulting from this thesis is the case, because for commuter train, train and tram the precision is zero. The average of the transportation modes that have a precision value is 57.16%. Bicylce and walk have lower precision values in this thesis compared to the paper, for bus they are similar and the precision for car is higher in this thesis.
For the accelerometer sensor, the classification results in an accuracy of 88.71% (tab. 5.14), which is lower than the values found in comparable studies in the literature. Shafique & Hato (2014) obtain an overall accuracy of 99.8% and Lu et al. (2018) an overall accuracy of 97.33%. Possible reasons for the discrepancy of accuracies are discussed in the following.
Shafique & Hato (2014) have more data available for the study, with over 3000 recorded trips, classifying them into only four different modes of transportation. Active modes to be classified are walking and bicycle, and for passive modes, only car and train are selected. The higher amount of data and the fewer transportation modes, reducing the possibility of misclassification, might explain the difference of accuracy. Lu et al. (2018) use only 30 hours of collected data by 10 different individuals and can reliably distinguish between five different transportation modes. A range of six seconds is used to compute accelerometer features for transportation mode detection. In comparison, for this thesis, over 150 hours of data are available, but the features are extracted from longer time windows, reducing the number of features of the data. This different approach could explain the reduction of accuracy for the method in this thesis.
When using both sensors, the highest overall accuracy of 91.77% is achieved (tab. 5.14), looking at the different transportation modes, an average precision of 84.69% can be calculated. From literature introduced in Chapter 2, Ellis et al. (2014) achieve a cross-validated accuracy of 89.8% and Martin et al. (2017) of 96.8%. The approach used for this thesis is based on Ellis et al. (2014), where features from the GPS and accelerometer are extracted from one-minute time windows of the data and classified into five different activities, including sitting and standing. Aside from the types of transportation modes,

the approaches are very similar, even the amount of data of 150 hours is nearly equal. Using the RF algorithm and with a similar approach, they achieved an accuracy close to the one from the MOASIS trained classifier. Martin et al. (2017) have an accuracy of around 5% more than for the MOASIS trained classifier, explanations for these differences aside from the specifics regarding the classifier and features might be that they classified only unimodal trips.

The results of classification for the single transportation modes depending on the sensors are more deeply discussed using the classifier trained and tested with the MOASIS dataset (fig. 5.23). The correct predictions per transportation mode are compared using the F1 scores, summarising prediction and recall. For this classification case predictions made using only GPS features provide the lowest F1 scores for all modes of transportation. For the transportation modes commuter train, train and tram no correct predictions are made. Actual commuter train segments are predicted to be walking segments, train segments tram or walk segments and tram segments car or walk segments.
From the confusion matrix for the GPS (fig. 5.16), the two transportation modes that occur most often in the dataset (car and walk) are often predicted as other modes of transportation. Nevertheless, due to the composition of the data, car and walk obtain the highest F1 scores. As for the order of precisions per transportation mode, Stenneth et al. (2011) have the highest precision accuracy for stationary, followed by walking. Bus has a very similar precision accuracy as in this thesis, and the other modes have higher precision accuracies in the literature.
In the confusion matrix showing the precision for the accelerometer-trained classifier (fig. 5.17), the highest values are always found in the diagonal, describing correct predictions, with the lowest value being 0.7738 for tram and the highest value 0.9116 for walk. Feng & Timmermans (2013) show a similar order in transportation mode classification, with walk being one of the highest and tram the lowest transportation mode to be correctly predicted. Interestingly, for train and tram there is a small increase in the F1 score when using both GPS data and accelerometer data, even though no predictions are made when using GPS only. Towards the difference in sensor, the increase in prediction accuracy described by the F1 score per transportation mode increases strongly from using GPS to accelerometer features. This result is reflected in the overall accuracy and kappa values for the classifier trained and tested with the same dataset and confirm the above-mentioned statement, that the use of accelerometer increases the performance when using training and testing data from the same environment.

Different hierarchical classifications are tried to improve the overall classification result, as was shown in Subchapter 5.3.4 (tab. 5.9). For the case where the classifier is trained and tested using different datasets, the same can be observed as before; the classification results are best in using GPS features. The highest kappa value is achieved for the distinction of passive and active modes of transportation with GPS features, whereas the highest accuracy comes from the classifier classifying the active modes of transportation into their respective classes. For the case where the classifiers are trained and tested with the same dataset, both the facts that using accelerometer features increases the accuracy (Bedogni et al., 2012) and the validity of the result increases further when both sensors are used (Xia et al., 2014) are confirmed. For all classifiers, an increase in accuracy and kappa value can be observed from the use of features from one sensor to the combination of both (tab. 5.10). The best results are achieved for the classification of the dataset into active and passive modes of transportation, with a kappa value of 0.9036 when using both sensors. This classification is relevant when the focus of transportation mode analysis is part of a health analysis. For accelerometer- and GPS & accelerometer-trained classifiers, good accuracy and kappa value are achieved for the distinction of transportation modes into private or public. This is a topic that is interesting to be used in an environmental context. The classification into the four

modes of transportation has an accuracy of >82% for both the classifiers trained including accelerometer features. With kappa values of >0.75 for both classifiers, a reliable initial distinction between the similar public modes of transportation can be made.
A comparable classification procedure has been implemented by Hemminki et al. (2013), where the transportation mode classification has been performed in using three different classifiers. They performed the study using only accelerometer features. For the kinematic motion classifier, which distinguishes walking from all other modes of transportation, they achieved an accuracy of 99%. The motorised classifier which distinguishes between the different transportation modes, after the removal of the stationary periods, achieves a precision of around 80% to detect the modality correctly. The two directly comparable classifiers that are implemented in this thesis are the first classifier, which distinguishes between active and passive transportation and the fourth classifier assigning the public transportation data into the four modes. For the classifier trained and tested with the MOASIS data using accelerometer features, the accuracy of the first classifier is 92.7% and for the fourth 82.55% (tab. 5.10), with a precision of 0.8923 for the first classifier and a mean precision of 0.8243 for the fourth classifier. In comparing the values to those found in prior literature that distinguishes between active and passive transportation (Hemminki et al., 2013), the accuracy achieved with the kinematic motion classifier is higher than that achieved with the classifier applied in this thesis. A reason for this discrepancy can be that Hemminki et al. (2013) have only walking as an active transportation mode, whereas the MOASIS data also includes bicycle. The classifier applied in this thesis predicting the classes for the public modes of transportation achieves very similar results.

6.1.4 Summary

For the segmentation of the trajectories, it is difficult to rank the performance between the different classifiers precisely. The performance is determined in using the segmented trajectories in the classifiers, thus including the classifier performance and not only the validity of segmentation. For all tested classifiers, there is a strong increase in accuracy from using GPS features to accelerometer features (tab. 5.1). The highest results are achieved in using both sensors data, but the highest increase of 15-20%, comes from using accelerometer data as opposed to GPS data. The potential of the change-point segmentation method has been shown, but the highest results are achieved using a one minute fixed window size segmentation method.

The highest-ranking features for different classifiers trained by both datasets are relatively simple statistical features explaining the behaviour of highest or lowest speed/acceleration in a segment. Information describing the range of values or standard deviation is the most useful feature for transportation mode. For the accelerometer, summarising the three axes into the total acceleration is more valuable for classification, as it removes the influence of the sensor's orientation. When using both sensors in combination, the importance of features provided by each sensor varies greatly depending on the characteristics of the dataset. For the benchmark dataset, GPS features are ranked with the highest importance, whereas for the MOASIS dataset accelerometer features are ranked highest.

Out of the three tested classifiers, the best results are achieved by the RF classifier. The classifier trained with the benchmark data and applied on the MOASIS data has a reduction of accuracy of only 1%, between the classifier using GPS features and the one using combined features. However, the kappa value drops of around 0.11, showing that including accelerometer features worsen accuracy for this scenario (tab. 5.3). When training and testing the classifier with data from the same dataset, there is a strong accuracy increase of 20% comparing GPS and accelerometer data. The kappa value is increased by 0.4.

When using the accelerometer data, the addition of GPS features leads to a smaller increase of around 3% of accuracy and 0.07 for the kappa value (fig. 5.13 and tab. 5.4). The increase in the results accuracy can be explained when looking at the classification into the seven transportation modes; for the GPS, three public transportation modes are not predicted.

Following the discussion in this chapter, the research question **"What is the classification accuracy variation for the same trajectories using only one sensor (GPS or accelerometer) or a combination of both sensors as input for the classification?"** can be answered in the following.
The importance of the sensors used depends on the characteristics of the classifiers and data. When using a classifier trained and tested with a different dataset, the best results are achieved in using only GPS data. On the other hand, for a classifier trained with the same dataset, the accuracy increases strongly (20%) from using GPS to using accelerometer features and when using both features, the increase from the accelerometer to both sensors is small. For the MOASIS data where there is no possibility for manual labelling, the best method is to use the benchmark-trained classifier with GPS features extracted from one-minute long segments. Due to the coverage of data by the GPS sensor, this is not a good option. If the possibility occurs for manual labelling and training the classifiers with enough data from the same environment, the results of the classification using the accelerometer provides useful information about the transportation modes, which would introduce the possibility of being able to classify more data as no GPS coverage is necessary.

6.2 Research Question 2 - classification of challenging transportation modes

The remaining problem of misclassification due to the similarity in behaviour between different modes of transportation is discussed in Chapter 2. Many studies that achieve very high results in transportation mode detection summarise motorised transportation modes into one class. Huss et al. (2014) observe an increase in kappa value up to 0.9 when combining those transportation modes using GPS features. Xia et al. (2014) combine the modes into the mode motorised to classify their segments. In doing so and using both GPS and accelerometer features, a very high accuracy of 96.13% has been achieved. As the idea of this thesis is to classify the most frequently occurring transportation modes reliably, it is not an option to summarise all motorised transportation modes as one type. To achieve a reliable classification of those similar transportation modes, specific features are selected that have few correlations with velocity. They are introduced in Chapter 4.3 and consist of the heading change rate and the stop rate for speed measurements and the count of the crossing of zero acceleration for accelerometer records in the three axes. In the following, their influence in the classifiers and the results of classification focusing on the similar modes of transportation bus, car, commuter train, train and tram are discussed.

The information of the experienced heading change to a segment is added to facilitate the distinction between car and public transportation. The variable importance with respect to the classifier trained with the benchmark dataset (fig. 5.12), show the range of the turning angle as the 11^{th} most important feature for distinguishing between private and public transportation. In addition, the maximum turning angle experienced in a segment is in the middle of the variable importance ranking (fig. 8.6, see Appendix 1.7). For the classifier trained with the MOASIS data, the turning angle features have no importance and are at the bottom of the variable ranking. This can explain the difference of environment in which a dataset has been gathered; in the urban environment of the benchmark dataset, the assumption about higher turning angles for cars is more valid than in a rural environment where the behaviour of public transportation is different.

Towards the importance of the specially selected features in the overall classification, generally, the selected features rank lower than others. For the GPS sensor, features of high importance describe the characteristics maximum and range of speed, which can be more generally applied to all modes. In terms of specific variables (fig. 8.4, see Appendix 1.5), nine out of the ten are the lowest-ranking variables for the classifier using data from MOASIS. Looking at the GPS trained classifiers for both datasets, the selected features rank in the lowest third. The specially selected features describe features that can be used to characterise more specific motorised transportation modes.
For the accelerometer-trained classifier with the MOASIS dataset, the three specially selected features are those with the lowest importance out of all, whereas for the benchmark trained accelerometer classifier, the count of zero crossings in the y-axis ranks in the middle. This difference could be a result of the precise instructions on how the sensor is worn for the recording of the benchmark dataset.

Looking at the classification results for the classifier trained with the benchmark dataset using GPS features, many of the predicted challenging modes of transportation are classified as car and walk. As there are many recorded car segments, the result of this transportation mode is higher. For the F1 score of the remaining four challenging transportation modes, all modes of transportation have low F1 scores of <0.2 (tab. 5.5). For the classifier trained and tested with MOASIS data, the results for the GPS classifier perform poorly on the challenging transportation modes (fig. 5.16). No predictions are made for commuter train, train and tram segments, which are all classified as walking.
For the classifier trained and tested with accelerometer features from different datasets, there are no predictions for bus and train segments; of the remaining two, both public transportation modes commuter train and tram have low F1 scores (fig. 5.22). When training and testing with the same dataset, the results are better, as the F1 score for bus is 0.4, and for the other challenging transportation modes, it is >0.68 (tab. 5.8). This shows that the approach to use only accelerometer features is valid for transportation mode detection.
When using both sensors combined, training and testing with different datasets results in no predictions for bus and train, with a low remaining F1 score for commuter train and tram. In using training and testing data from the same environment, the use of both sensors increases the results from the accelerometer sensor approach slightly.

The results of Classifier 4 of the hierarchical classification approach are discussed to have an isolated look at the misclassification between the four challenging transportation modes. For the classifier trained using the GPS features of the benchmark data and applied on the MOASIS data, the accuracy is 33.61%, with a kappa value of 0.1324. In analysing the confusion matrix, sources of misclassification are analysed. The highest F1 score is for bus, followed by similar values for commuter train and tram. The transportation mode with the lowest F1 score is train (tab. 5.11). Due to the homogenous environment where the transportation modes are collected, it is difficult to distinguish them, due to a similarity in features. This misclassification can be reduced by including training data of public transportation collected outside of the urban environment, such as is the case for MOASIS. The occurring misclassification between train and commuter train can be explained with the similarity in features of these transportation modes. In addition, there is only a little amount of data available, as short segments are recorded with this transportation mode due to signal outage. In the geographical range where the benchmark dataset has been recorded, no clear distinction is possible as no long-distance trains are included in the dataset;

the results are even worse when using accelerometer features. With a total accuracy of 23.77% and a negative kappa value, the results of this classification are unusable. It is difficult to determine a specific pattern in the misclassifications as every false combination occurs using the accelerometer features. This is expectable, as these public transportation modes in the urban environment have very similar acceleration / deceleration patterns. In addition, these measurements are not influenced by the individual recording them; thus the records for these modes are quite homogenous.

To see what misclassification still occurs when the performance of the classifier is good, the results of testing and training with the data from the same environment and using both GPS & accelerometer features are discussed (tab. 5.13). Here the accuracy is 84.91%, with a kappa value of 0.7895. For the confusion matrix, the good results of classifying the modes of transportation are visible by the highest values being in the diagonal (fig. 5.20). Looking at the F1 score, the transportation mode with the lowest value of 0.7742 is commuter train, and the highest value is achieved for tram segments with a value of 0.9012. The general misclassification is small, but in the following, the most frequently occurring misclassifications per class are listed:

- Bus segments: Predicted as tram segments
- Commuter train segments: Predicted as train, tram and bus segments
- Train segments: Predicted as bus and tram segments
- Tram segments: Predicted as bus and commuter train segments

Due to the inclusion of transportation modes on a larger geographical scale and in different types of public transportation environments, the misclassification between bus and tram segments is strongly reduced. This reduction comes from the data and not from the features selected. When looking at the precision values, the strongest misclassification occurs between commuter train and train segments, which is understandable due to their similarity in features, especially when there are not many segments describing each transportation mode.
There are still improvements to be done for the reliable classification of those transportation modes, but it shows that GPS and accelerometer data from the same environment used in a RF classifier provide a reliable first classification of the four transportation modes, assuming that only those modes occur in the dataset.

Stenneth et al. (2011) achieve precision accuracies of 69.8 for train and 56.5 for bus segments for the RF classifier, which uses GPS features. Comparing the values to the isolated GPS sensor classifiers applied in this thesis, which predict only public transportation modes, the precision accuracies are lower. This can come from the different classification procedure, the classifier used in Stenneth et al. (2011) includes all transportation modes. The precision for the classifier trained and tested with different data, (tab. 5.11), is 25 for train and 31.62 for bus segments. For the classifier trained and tested with the MOASIS data (tab. 8.3, see Appendix 1.9), the precision is 62.5 for train and 59.18 for bus segments, which is very similar to the ones seen in Stenneth et al. (2011). When accelerometer features are added, the precisions increase by around 20.
Feng & Timmermans (2013) list the confusion matrices for the result of the classification of similar transportation modes using Bayesian Belief Network. For GPS the precision for bus is 78, for train 89, and for tram 83, when using accelerometer these values are increased with a precision of 87 for bus, 98 for tram and 58 for train. When both sensors are used for the classification, bus has a precision of 98, train of 83 and 99 for tram. It is confirmed in these values that there is a trend in increasing prediction values

for transportation modes using accelerometer features. Due to the very low kappa value seen in this thesis (tab. 5.9), the results of the classifier trained and tested with a different dataset is not further discussed. For the classifier trained and tested with the MOASIS data using accelerometer features (tab. 5.13), the ranking of precision show the same trend, but with lower values. The precision is highest for tram (90.12), followed by bus (83.87) and train (81.58).

Different studies describe the inclusion of additional GIS data, such as open street map data, to improve the results of transportation mode classification. The inclusion of such data not only solves ambiguity between different similar modes of transportation and gives additional information about the recorded trajectories, but it can also be used to reduce the error of data with gaps. Including this information, a classification accuracy of 91.6% has been obtained (Biljecki et al., 2013). Going further in this direction, the inclusion of additional information about the transportation network, such as the real-time public transportation locations and information about the routes increases the accuracy significantly (Stenneth et al., 2011).

6.2.1 Summary

Different classification approaches are tested to try and improve the transportation mode detection of challenging modes. For one, specific features are selected from both sensors that should help distinguish modes of transportation without the influence of traffic. Overall, the use of these features does not prove helpful, as seen in the variable importance scores of the different classifiers. The only useful feature is the range of turning in a segment, which proves important for the distinction of private and public transportation trained with the benchmark dataset. This shows the influence that the characteristics of the training dataset have. For the benchmark dataset, which was mostly collected in an urban environment some features are more important than for the MOASIS dataset, which was collected over a much larger geographical area. Results for the distinction of the four challenging modes of transportation, bus, commuter train, train and tram in the classifier containing all modes of transportation provide low F1 scores for the modes of transportation when training and testing data come from a different dataset. The answer to the question **"How can misclassifications between transportation modes which have a similar signal in an urban context be reduced (car, bus, tram)?"** must be discussed from different points of view. Regarding the introduction of specific variables in helping the classification, the heading change statistics, stop time and count of zero crossings proposed in this thesis do not facilitate the differentiation, as they are ranked as variables with low importance by the classifiers. Towards the reliable classification, there are poor results when using a classifier trained by data collected separately from the testing data, especially for the public transportation modes, which are often not predicted. For the classifier trained and tested with data from the same environment, the results for GPS features are bad, for the accelerometer features they are better. The highest results are achieved when a hierarchical classifier distinguishing between the public transportation modes is trained with data from the same environment as the testing data, as only those modes are found in the data. This includes the use of both sensors and assumes that previous classifiers correctly predicted the other modes of transportation. Thus, to improve the classification of these modes of transportation, it is advised to include further GIS data, such as information about the public transportation stops or network information.

6.3 Critical Evaluation

In this subchapter, some aspects of this thesis and the data used are critically discussed. One obstacle is that the data used from the MOASIS study, needs to be manually selected and labelled, and gaps due to signal outage removed. In addition, the frequency of the collected data is more problematic than expected, with many records having only one entry per minute. Following this and the labelling interface used, this results in the same sampling frequency of accelerometer data as for the GPS data. This might additionally introduce an error in the classification using the accelerometer features as it means transferring the benchmark trained classifier with a much higher frequency of the accelerometer records to the MOASIS data with lower accelerometer frequency. Regarding this problematic, it was found that where the testing data rate is lower than the training data rate, as seen in this thesis, the accuracy should decrease only slightly. It has also been shown, that the sampling at a high frequency does not improve the performance significantly, this information is important towards minimising the energy consumption of the sensor while collecting the transportation data (Bedogni et al., 2012). Besides, the manual labelling may have introduced errors. Biljecki et al. (2013) state that after observing the difference between manual and automated classification, approximately the same number of mistakes can be found. These mistakes include different modes of transportation classified in one segment or the neglecting of short segments.

Additionally, during labelling, transportation modes occurring without being part of the modes from the benchmark dataset were removed. For the selected part of the MOASIS data, this includes modes of transportation such as cable car, ship and moped, as well as kayaking. These methods are not detected by the classifiers due to the lack of ground truth data but could be considered for future research.

In future, before applying the discussed methods on the dataset, a large amount of pre-processing is necessary, such as the removal of gaps and stationary periods. As the discrepancy between the two datasets is large, one solution for the future training of classifiers that can be used to classify different data is to mix the qualitatively perfect benchmark data set, which has a restricted geographical dispersion, with the MOASIS data which records real-life conditions in more diverse environments. To build a classifier that can be applied to a random dataset, it is essential to include data about mobility from different data collections and environments.

For the applied classifiers, one of the difficulties is dealing with strongly unbalanced training data. Within this thesis, the unbalance was handled in up-sampling segments of transportation modes with lower occurrences. Interestingly, the distribution of both datasets regarding the modes of transportation is very similar. However, the results of the classifier show that often segments are predicted to be car and walk, which are the two transportation modes that occur most frequently. To reduce this influence, a stronger balance between the transportation modes is necessary for the development of a better benchmark dataset.

Towards the transferring of a classifier trained with a different dataset than it will be applied on, the difference of characteristics of the two datasets is large. Initially, differences were suspected regarding the influence of the age of the individuals, which can be disregarded for macro-scale mobility analysis, as the most significant influence comes from the environment. The benchmark dataset has been collected mostly in an urban environment, which is not suitable to be used for data classification of different environments. Even though it has been collected in the real world, the individuals followed strict instructions, resulting in data collected in a controlled way which is almost too perfect. The real-life dataset also has modes of mobility that are difficult to define, for example, the transferring between different transportation modes, which can be waiting at a bus stop, without necessarily being stationary. Such segments must be removed before the analysis. As for the transportation modes that are selected in the benchmark dataset, the most important ones are undoubtedly present. An improvement would be to combine train and commuter train into one class, as this distinction introduces unnecessary errors. If the travel behaviour regarding distance travelled needs to be analysed, it could be done with the GPS records. Otherwise, there seems to be no additional information gained in distinguishing between different train types.

7 Conclusion

7.1 Summary

The major findings of this thesis are that even in testing different classifiers and different sensor combinations, the transfer of a classifier trained on one dataset to classify a different dataset remains challenging. Generally, the RF algorithm is an adequate classifier for transportation mode detection, due to its ability to build multiple classifiers using different features of the data, with the plurality vote, the final prediction is obtained from these classifiers.
When using a classifier trained with different transportation data, the overall classification results are poor. The classification using GPS features provides the highest accuracies and similar results are achieved using both sensors. When using accelerometer features for classification, the results are not usable, following the classification of all segments as the transportation mode walk.
Towards the sensor coverage of the MOASIS data, it must be mentioned that the classifier trained with the accelerometer features achieve a higher value than using only GPS features. When both sensors are available, the results are slightly improved compared to accelerometer classification. The trend of classification can be compared to literature; if the training and testing data come from the same data collection, the best classifiers use both GPS and accelerometer features. It was proven that in using simple features, the characteristics of the different transportation modes could be well captured, providing a good overview of the transportation mode detection.

Towards the correct classification of transportation modes that have similar features, the best classification results are achieved when using both sensors and a classifier that has been trained and tested with data from the same environment. For classifiers that only focus on distinguishing between public and private transportation and classify the public transportation modes, reliable results are achieved when only data with such transportation modes is included. Again, the best results are achieved when using data from the same environment, utilising features from both sensors. In that case, the precision value for distinguishing public and private transportation is 0.77 and for classifying the public transportation the F1 score for the four classes are >0.77. These values show that the applied methods are a good start, but the specially selected features are useless as they are ranked low in importance for the classifiers. Following the literature, additional GIS information, such as information about public transportation stop locations or the type of road, would help increase the validity of classification results.

The analyses performed within this thesis generally confirmed the results discussed by prior literature for the different sensor usage when training and testing the classifier with data from the same dataset. Specifically building a universal classifier that can be used for transportation mode classification of non-labelled data is not discussed in the literature. The results obtained from one classifier and applied onto different data reduce the results of classification substantially, especially for the accelerometer sensor.

7.2 Future work

For future research, a more in-depth analysis of how to build a classifier, that can be used to classify different datasets containing trajectories recorded with GPS and accelerometer into different transportation modes, must be examined. After collecting real-life data, the raw measurements must be pre-processed to extract macro-scale mobility trajectories automatically. This implies removing micro-scale mobility, stationary periods, as well as outliers. When this is achieved, the idea of the benchmark dataset needs to be expanded. In using only data collected in the real world, but in a controlled way, the dataset is not suited for training or testing classifiers, as it does not reflect spontaneous real-life behaviour. In order to get a steadier dataset that can either be used to train a classifier and apply it on unlabelled data or to comparatively test classifiers which have been trained using different datasets, the expansion of the benchmark dataset is necessary. In including data collected by different individuals and during different studies in different environments, the dataset gets more robust. Ellis et al. (2014) state that there is a homogeneity in recording a prescribed trip, regardless of the person recording it. Thus, they state that using data from different environments is more important to validate results than to have data from different individuals.

The difficulty in combining data from different studies for building a benchmark dataset is to be able to control and guarantee the quality of the data. For the benchmark dataset, the quality is high because the individual data collection was controlled with specific instructions. A strength of the benchmark and MOASIS data is that they have been collected both by the same device, reducing a possible influence of the device onto the results. Nevertheless, the frequency by which the data is collected varies greatly. For this thesis, the minimal frequency was selected to be at least 0.5Hz, but there is a broad range with few entries having a frequency of up to 8Hz. The influence of such different sampling rates onto classifiers and their influence on the results must be studied. Not only is this necessary to create a better benchmark dataset, but it is essential to include trajectories recorded in different environments. The benchmark dataset has recorded most of the public transportation segments only in urban areas, whereas the MOASIS data has a larger geographical area of data collection, including rural and urban environments. The creation of a robust classifier, which can classify transportation modes from different datasets reliably, should include such diverse data for it to be applied on a larger scale, such as a country or an area where the main transportation modes are the same.

Literature

Allahbakhshi, H. & Weibel, R. (2017). Using multi-sensor tracking data to analyze the mobility and activity behavior of older adults Table of Contents. In *4th AGILE PhD School* (pp. 1–32).

Allahbakhshi, H. & Weibel, R. (2018). A Study Design for Physical Activity Reference Data Collection Using GPS and Accelerometer. In *21st International Conference on Geographic Information Science (AGILE 2018)* (pp. 1–6).

Bedogni, L., Di Felice, M., & Bononi, L. (2012). By train or by car? Detecting the user's motion type through smartphone sensors data. In *IFIP Wireless Days* (pp. 1–6).: IEEE.

Bereuter, P., Fillekes, M. P., & Weibel, R. (2016). Analyzing GPS and accelerometer data in the study of the Mobility, Activity and Social Interaction of older adults. In *Visually-supported Computational Movement Analysis (VCMA 2016)*.

Bereuter, P. & Weibel, R. (2016). Ambulatory assessment to study mobility and activity patterns in healthy aging research. *13th International Conference on Location-Based Services (LBS 2016)*, (November), 124–128.

Bhowan, U., Johnston, M., Zhang, M., & Yao, X. (2013). Evolving diverse ensembles using genetic programming for classification with unbalanced data. *IEEE Transactions on Evolutionary Computation*, 17(3), 368–386.

Biljecki, F., Ledoux, H., & van Oosterom, P. (2013). Transportation mode-based segmentation and classification of movement trajectories. *International Journal of Geographical Information Science*, 27(2), 385–407.

Boulesteix, A. L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6), 493–507.

Brondeel, R., Pannier, B., & Chaix, B. (2015). Using GPS, GIS, and Accelerometer Data to Predict Transportation Modes. *Medicine and Science in Sports and Exercise*, 47(12), 2669–2675.

Buchin, M., Driemel, A., Kreveld, M. V., & Sacrist, V. (2011). Segmenting trajectories : A framework and algorithms using spatiotemporal criteria *. *Journal of Spatial Information Science*, 3, 33–63.

Bundesamt für Statistik (2018). Seniorinnen und Senioren. *https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/alterung/senioren.assetdetail.6666880.html*[Accessed: 01.05.2019].

Burkhard, O. (2017). Movement data Annotation. *https://oliver-burkhard.shinyapps.io/17fs/movement/annotation/interface/* [Accessed 23.04.2019].

Byon, Y. J., Abdulhai, B., & Shalaby, A. (2009). Real-time transportation mode detection via tracking global positioning system mobile devices. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 13(4), 161–170.

Byon, Y. J. & Liang, S. (2014). Real-time transportation mode detection using smartphones and artificial neural networks: Performance comparisons between smartphones and conventional global positioning system sensors. *Journal of Intelligent Transportation Systems*, 18(3), 264–272.

Das, R. & Winter, S. (2016). Detecting Urban Transport Modes Using a Hybrid Knowledge Driven Framework from GPS Trajectory. *ISPRS International Journal of Geo-Information*, 5(207).

Ellis, K., Staudenmayer, J., Kerr, J., Marshall, S., Lanckriet, G., & Godbole, S. (2014). Identifying Active Travel Behaviors in Challenging Environments Using GPS, Accelerometers, and Machine Learning Algorithms. *Frontiers in Public Health*, 2(April), 1–8.

Fang, S. H., Fei, Y. X., Xu, Z., & Tsao, Y. (2017). Learning Transportation Modes from Smartphone Sensors Based on Deep Neural Network. *IEEE Sensors Journal*, 17(18), 6111–6118.

Feng, T. & Timmermans, H. J. (2013). Transportation mode recognition using GPS and accelerometer data. *Transportation Research Part C: Emerging Technologies*, 37, 118–130.

Fillekes, M. P., Röcke, C., Katana, M., & Weibel, R. (2019). Self-reported versus GPS-derived indicators of daily mobility in a sample of healthy older adults. *Social Science and Medicine*, 220, 193–202.

Gardner, M. W. & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14/15), 2627–2636.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3–42.

Gong, H., Chen, C., Bialostozky, E., & Lawson, C. T. (2012). A GPS/GIS method for travel mode detection in New York City. *Computers, Environment and Urban Systems*, 36, 131–139.

Gonzalez, P. A., Weinstein, J. S., Barbeau, S. J., Labrador, M. A., Winters, P. L., Georggi, N. L., & Perez, R. (2008). Automating mode detection using neural networks and assisted GPS data collected using GPS-enabled mobile phones. In *World Congress on Intelligent Transportation Systems*.

Guo, T., Yan, Z., & Aberer, K. (2012). An adaptive approach for online segmentation of multi-dimensional mobile data. In *MobiDE '12 Proceedings of the Eleventh ACM International Workshop on Data Engineering for Wireless and Mobile Access* (pp. 7–14).

Hansen, B. H., Kolle, E., Dyrstad, S. M., Holme, I., & Anderssen, S. A. (2012). Accelerometer-determined physical activity in adults and older people. *Medicine and Science in Sports and Exercise*, 44(2), 266–272.

Hemminki, S., Nurmi, P., & Tarkoma, S. (2013). Accelerometer-based transportation mode detection on smartphones. *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems - SenSys '13*, (pp. 1–14).

Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2010). A Practical Guide to Support Vector Classification.

Huss, A., Beekhuizen, J., Kromhout, H., & Vermeulen, R. (2014). Using GPS-derived speed patterns for recognition of transport modes in adults. *International Journal of Health Geographics*, 13(1), 40.

Isler, V. (2018). *Building and Using a Reference Data Set for Transportation Mode Detection*. PhD thesis.

Jahangiri, A. & Rakha, H. A. (2014). Developing a Support Vector Machine (SVM) Classifier for Transportation Mode Identification by Using ... In *Transportation Research Record*.

Jahangiri, A. & Rakha, H. A. (2015). Applying Machine Learning Techniques to Transportation Mode Recognition Using Mobile Phone Sensor Data. *IEEE Transactions on Intelligent Transportation Systems*, 16(5), 2406–2417.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer New York Heidelberg Dordrecht London.

Kaghyan, S. & Sarukhanyan, H. (2013). Accelerometer and GPS sensor combination based system for human activity recognition. In *CSIT 2013 - 9th International Conference on Computer Science and Information Technologies, Revised Selected Papers* (pp. 1–9).: IEEE.

Kuhn, M. (2019). Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan and Tyler Hunt. (.

Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their Applications*, 83, 79–102.

Lin, Y., Lee, Y., & Whaba, G. (2002). Support Vector Machines for Classification in Nonstandard Situations. *Machine Learning*, (46), 191–202.

Lu, D. N., Nguyen, D. N., Nguyen, T. H., & Nguyen, H. N. (2018). Vehicle mode and driving activity detection based on analyzing sensor data of smartphones. *Sensors (Switzerland)*, 18(4), 1–25.

Mäenpää, H., Lobov, A., & Martinez Lastra, J. L. (2017). Travel mode estimation for multi-modal journey planner. *Transportation Research Part C: Emerging Technologies*, 82, 273–289.

Marconi, D. (2008). *Auswirkungen des demografischen Wandels auf die Mobilität. Das Verkehrsverhalten der heutigen und der künftigen Senioren*. Technical report.

Martin, B. D., Addona, V., Wolfson, J., Adomavicius, G., & Fan, Y. (2017). Methods for real-time prediction of the mode of travel using smartphone-based GPS and accelerometer data. *Sensors (Switzerland)*, 17(9), 1–20.

McLean, D. J. & Skowron Volponi, M. A. (2018). trajr: An R package for characterisation of animal trajectories. Ethology, 124(6), 440-448.

Mollenkopf, Heidrun, Marcellini, F., Ruoppila, I., Flaschenträger, P., Gagliardi, C., & Spazzafumo, L. (1997). Outdoor mobility and social relationships of elderly people. *Archives of Gerontology and Geriatrics*, 24(3), 295–310.

Murphy, S. L. (2009). Review of physical activity measurement using accelerometers in older adults: Considerations for research design and conduct.

Nguyen, D. M. T., Lecoultre, V., Sunami, Y., & Schutz, Y. (2013). Assessment of Physical Activity and Energy Expenditure by GPS Combined with Accelerometry in Real-Life Conditions. *Journal of Physical Activity and Health*, 10, 880–888.

Nikolic, M. & Bierlaire, M. (2017). Review of transportation mode detection approaches based on smartphone data. *17th Swiss Transport Research Conference*, (May).

Patin, R., Etienne, M.-p., Lebarbier, E., & Benhamou, S. (2019). segclust2d: Bivariate Segmentation/Clustering Methods and Tools. R package version 0.2.0.

Patin, R., Etienne, M.-p., Lebarbier, E., Chamaillé-Jammes, S., & Benhamou, S. (2018). *Identifying stationary phases in multivariate time-series for highlighting behavioural modes and home range settlements*. Technical report.

Peixeiro, M. (2019). Get Started with Support Vector Machines (SVM).

Picard, F., Robin, S., Lebarbier, E., Daudin, J., Ina, U. M. R., & Inra, P.-g. E. (2007). A Segmentation / Clustering Model for the Analysis of Array CGH Data. *Biometrics*, 63(September), 758–766.

Pires, I. M., Garcia, N. M., Pombo, N., Flórez-Revuelta, F., & Spinsante, S. (2017). Pattern Recognition Techniques for the Identification of Activities of Daily Living using Mobile Device Accelerometer. *PeerJ PrePrints 2017*.

Prelipcean, A. C., Gidofalvi, G., & Susilo, Y. O. (2016). Measures of transport mode segmentation of trajectories. *International Journal of Geographical Information Science*, 30(9), 1763–1784.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., & Srivastava, M. (2010). Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks*, 6(2), 1–27.

Schrack, J. A., Cooper, R., Koster, A., Shiroma, E. J., Murabito, J. M., Rejeski, W. J., Ferrucci, L., & Harris, T. B. (2016). Assessing Daily Physical Activity in Older Adults: Unraveling the Complexity of Monitors, Measures, and Methods. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, 71(8), 1039–1048.

Shafique, M. A. & Hato, E. (2014). Use of acceleration data for transportation mode prediction. *Transportation*, 42(1), 163–188.

Stenneth, L., Wolfson, O., Yu, P. S., & Xu, B. (2011). Transportation mode detection using mobile phones and GIS information. In *ACM Sigspatial GIS'11* (pp.$\tilde{5}$4).

Varicchio, C. (2016). rowr: Row-Based Functions for R Objects. R package version 1.1.3.

Vilhelmson, B. (1999). Daily mobility and the use of time for different activities . The case of Sweden. *GeoJournal*, 48, 177–185.

Wang, S., Chen, C., & Ma, J. (2010). Accelerometer based transportation mode recognition on mobile phones. *APWCS 2010 - 2010 Asia-Pacific Conference on Wearable Computing Systems*, (pp. 44–46).

Widhalm, P., Nitsche, P., & Brändle, N. (2012). Transport Mode Detection with Realistic Smartphone Sensor Data. In *21st International Conference on Pattern Recognition* (pp. 573–576).

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data Mining. Practical Machine Learning Tools and Techniques*. Fourth edi edition.

Wu, L., Yang, B., & Jing, P. (2016). Travel mode detection based on GPS raw data collected by smartphones: A systematic review of the existing methodologies. *Information (Switzerland)*, 7(4).

Xia, H., Qiao, Y., Jian, J., & Chang, Y. (2014). Using Smart Phone Sensors to Detect Transportation Modes. *Sensors*, 14, 20843–20865.

Xiao, Z., Wang, Y., Fu, K., & Wu, F. (2017). Identifying Different Transportation Modes from Trajectory Data Using Tree-Based Ensemble Classifiers. *ISPRS International Journal of Geo-Information*, 6(57), 1–22.

Yu, M.-c., Yu, T., Wang, S.-C., Lin, C.-J., & Chang, E. Y. (2014). Big Data Small Footprint : The Design of A Low-Power Classifier for Detecting Transportation Modes Categories and Subject Descriptors. In *40th International Conference on Very Large Data Bases* (pp.$\tilde{1}$2).

Zheng, Y., Liu, L., Wang, L., & Xie, X. (2008). Learning transportation mode from raw gps data for geographic applications on the web. *Proceeding of the 17th international conference on World Wide Web - WWW '08*, (pp. 247).

Zheng, Y., Yukun, C., Quannan, L., Xing, X., & Wei-Ying, M. (2010). Understanding transportation modes based on GPS data for web applications. *ACM Transactions on the Web*, 4(1), 1–36.

8 Appendices

Appendix 1: Additional tables and figures for results chapter

1.1 Table including the classification results of the three different classifiers trained and tested with different sensor combinations. The accuracy and kappa-values are listed in the table, opposing the results of the two segmentation methods, which are the segmentation by time interval (minute) or segmentation by the labels.

Table 8.1: Accuracy and Kappa values for the classifiers trained and tested with the different sensors. Data segmented by minute-intervals or by labels is used for training and testing.

Classifier	GPS by Minute	Accelerometer by Minute	GPS + Accelerometer by Minute	GPS by Label	Accelerometer by Label	GPS + Accelerometer by Label
SVM						
Accuracy	0.6424	0.7534	0.8223	0.6358	0.7107	0.6377
Kappa	0.5308	0.3566	0.7596	0.5229	0.6288	0.4757
RF						
Accuracy	0.6689	0.7828	0.8571	0.5699	0.7397	0.7101
Kappa	0.5507	0.7213	0.8056	0.4554	0.6584	0.5643
NN						
Accuracy	0.6556	0.7051	0.7317	0.6218	0.686	0.6232
Kappa	0.5434	0.6304	0.6407	0.5301	0.5939	0.4745

1.2 Table with the names of features in the variable importance figures and the explanation. For the accelerometer features, at the end of the name the X, Y, Z or total indicates the axes on which it has been computer. Total is the total acceleration.

Table 8.2: Overview feature names in variable importance plots.

Name in figure	Feature
Max/ min	Maximum/ minimum of speed/acceleration
Mean / median	Mean/ median of speed/acceleration
Sd/ variance	Standard deviation / variance of speed/acceleration
Delta	Range of speed/acceleration
Ninetyfifth / Eightieth / Twentieth	95th / 80th / 20th percentile of speed/acceleration
interquartile	Interquartile range of speed/acceleration
RMS	Root Mean Square of speed/acceleration
Kurtosis / skewness	Kurtosis/skewness of speed/acceleration
Coefficient	Coefficient of variation of speed
turningAngleSum, maxTurningAngle, minTurningAngle, deltaTurningAngle	Heading change rate for GPS (sum, maximum, minimum and range)
stopTime	Time person stops in a segment
zeroCrossX/Y/Z	Zero-crossing of the accelerometer signal

1.3 Variable importance for the SVM and NN classifiers using the three different sensor combinations.

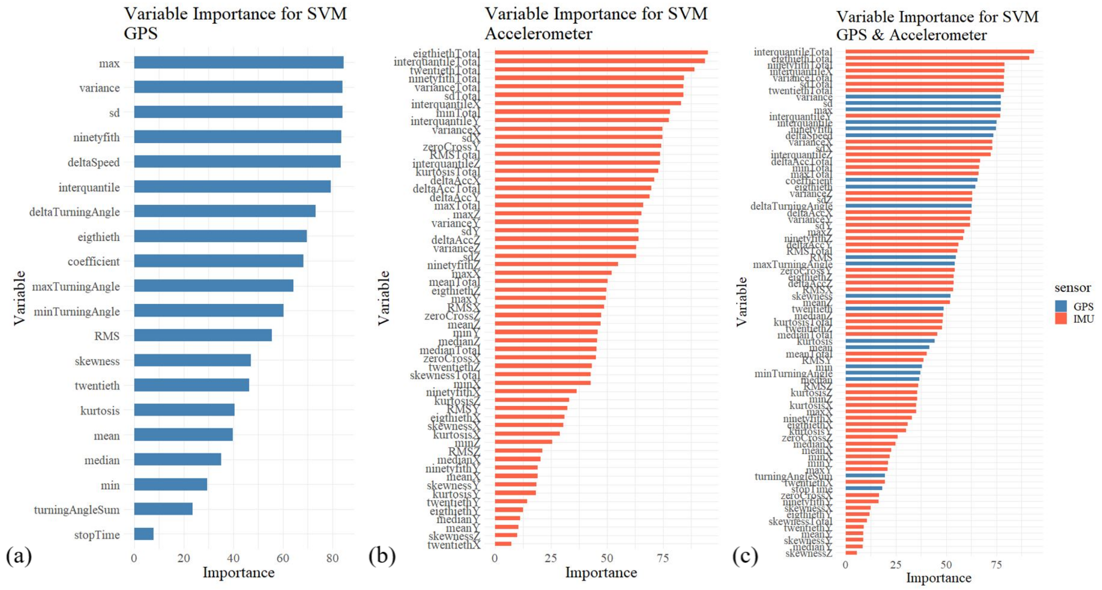

Figure 8.1: Variable importance for the SVM classifier trained with the benchmark dataset. In (a) the SVM is trained only with features from the GPS sensor, in (b) with features from the accelerometer sensor and in (c) with the combination both sensors.

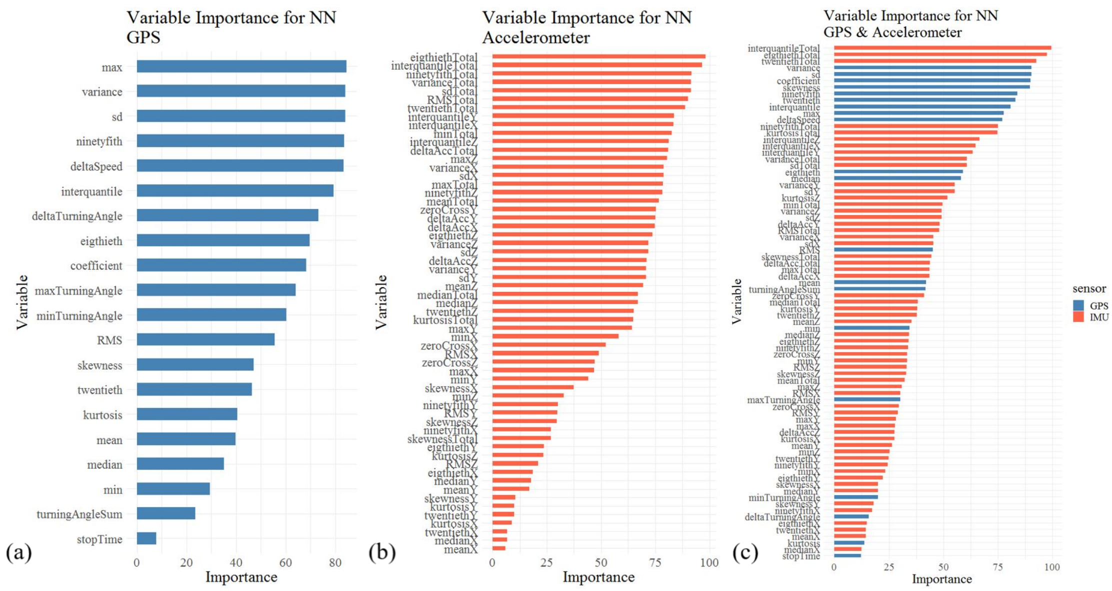

Figure 8.2: Variable importance for the NN classifier trained with the benchmark dataset. In (a) the NN is trained only with features from the GPS sensor, in (b) with features from the accelerometer sensor and in (c) in combining both sensors.

1.4 Variable importance for the classification of the four public transportation classes. All have been trained using only benchmark data from the four public transportation modes bus, commuter train, train and tram.

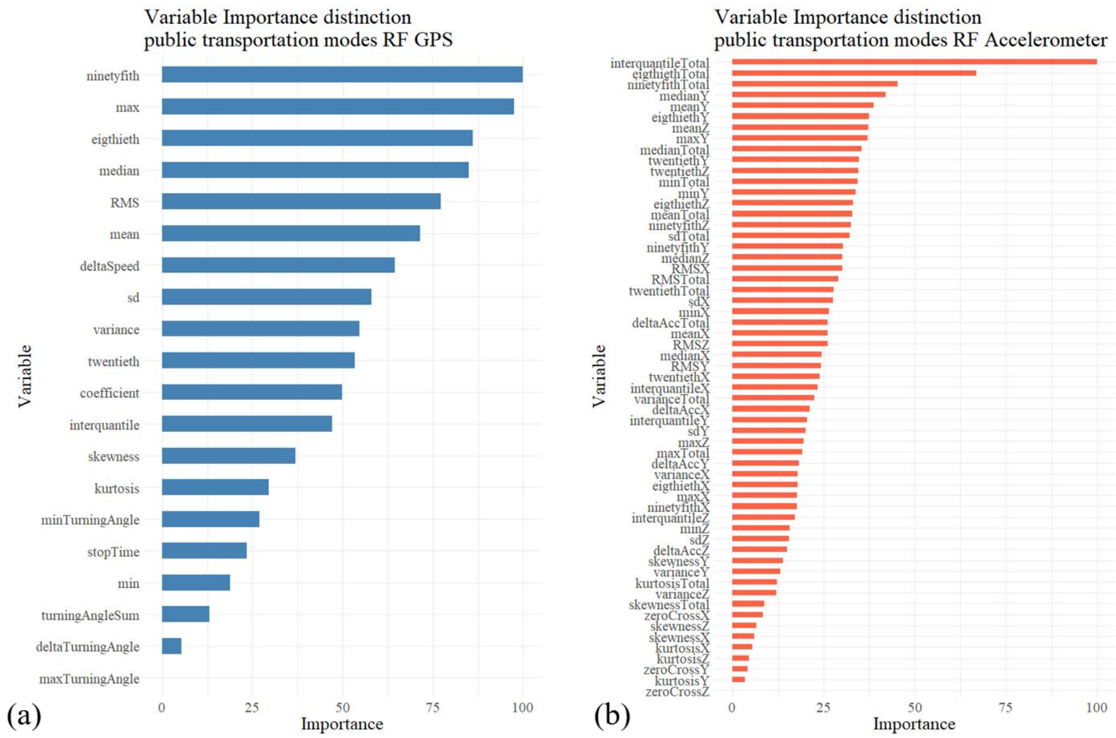

Figure 8.3: Variable importance for the classifier into the four public transportation modes using RF and the benchmark data. In (a) the variable importance for the GPS classifier is illustrated and in (b) of the accelerometer classifier.

1.5 The variable importance for the classifiers trained with the MOASIS data and the RF classifier.

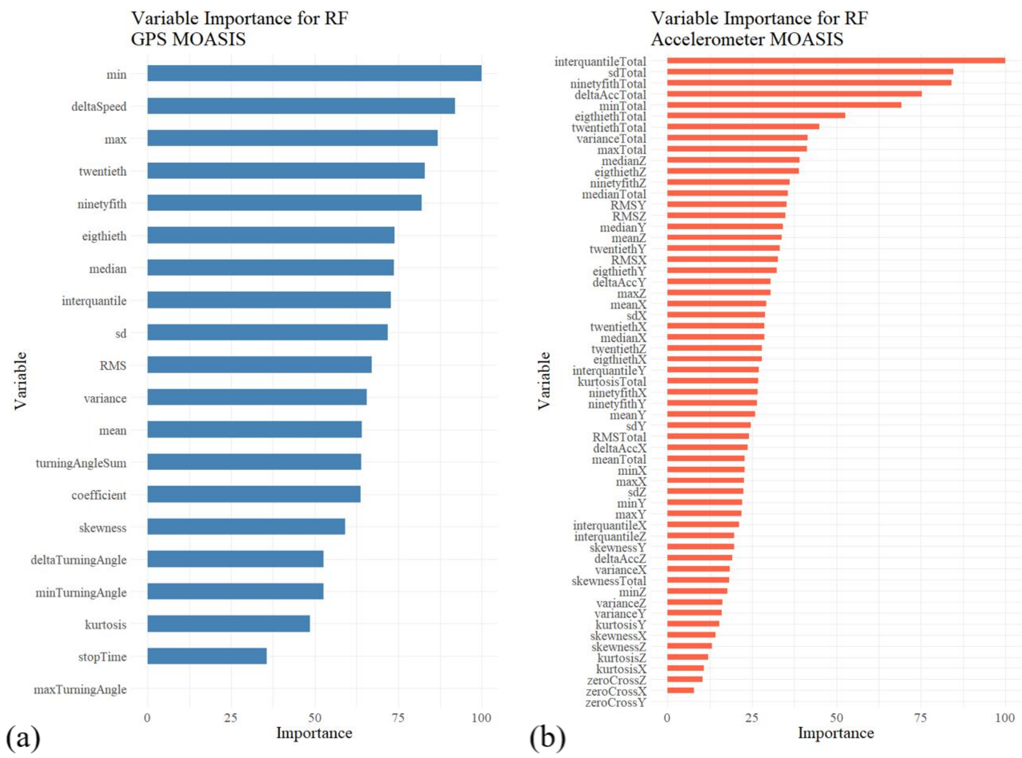

Figure 8.4: Variable importance for the RF classifier trained with the MOASIS data, in (a) using GPS sensor features and in (b) using accelerometer features.

1.6 Importance of the specially selected variables that were selected to improve the classification of the challenging transportation modes for the GPS and accelerometer measurements of the benchmark dataset.

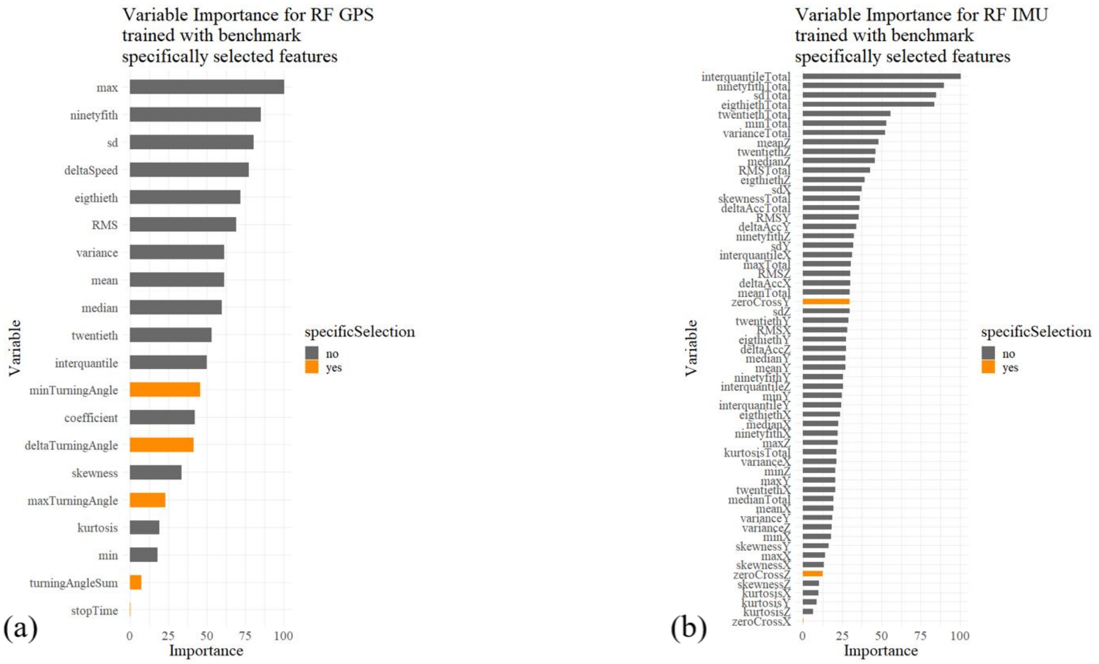

Figure 8.5: Illustrations of variable importance of the RF classifier trained with the benchmark dataset. The specially selected variables are highlighted in orange. In (a), features from the GPS sensor can be seen and in (b) from the Accelerometer sensor.

1.7 Variable importance of the specially selected variables for the RF classifier trained with the MOASIS dataset.

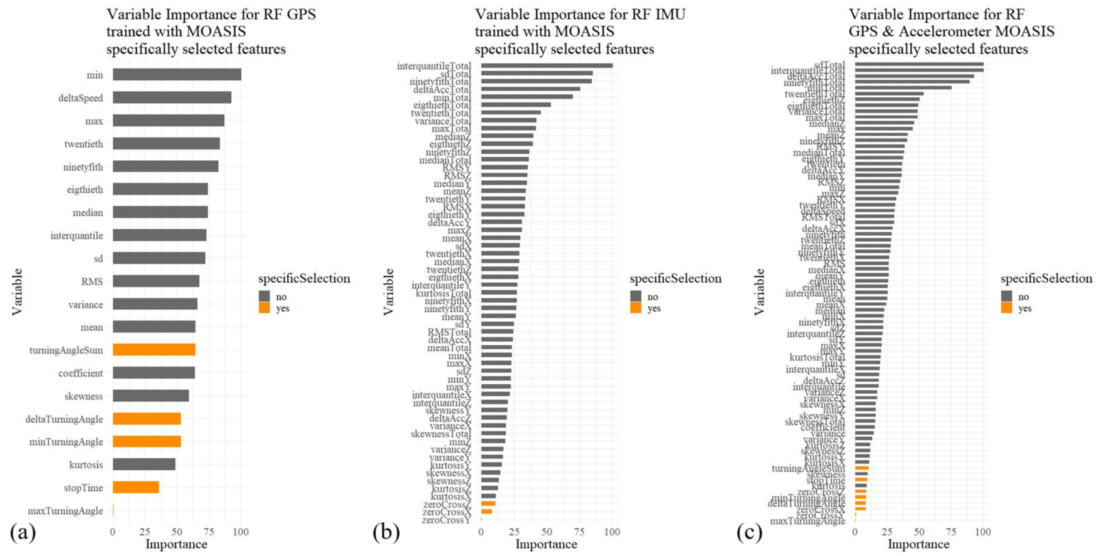

Figure 8.6: Illustrations of variable importance of the RF classifier trained with the MOASIS dataset. The specially selected variables are highlighted in orange. In (a), features from the GPS sensor can be seen, in (b) from the accelerometer sensor and in (c) from using both sensors.

1.8 Highlighting the specially selected variables in the variable importance list for classifiers trained with the MOASIS data. The classifiers distinguish between private and public transportation and classify public transportation into its respective modes.

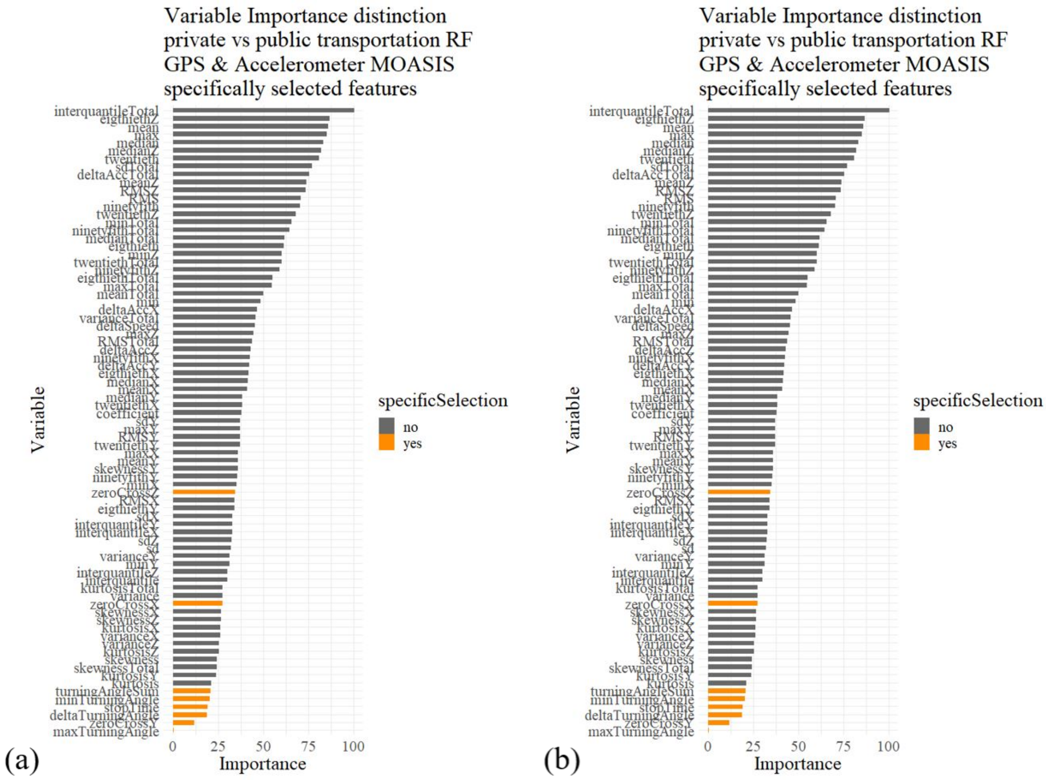

Figure 8.7: Variable importance for the RF classifier trained with the MOASIS dataset. The specially selected features to distinguish challenging transportation modes are highlighted in orange. In (a), the feature importance of the classifier distinguishing between private and public transportation is shown and in (b) the classifier determining the four public transportation modes. Both sensors were used for this analysis.

1.9 Precision, Recall and F1 score of the classification into the four public transportation modes for the classifier trained and tested with MOASIS using either GPS or Accelerometer features.

Table 8.3: Precision, Recall and F1 score per transportation modes for the RF classifier distinguishing between the four public transportation modes. The classifier is trained and tested on the MOASIS dataset using GPS features.

Mode (using GPS data)	Precision	Recall	F1
Bus	0.5918	0.5273	0.5577
Commuter Train	0.4706	0.4324	0.4507
Train	0.625	0.4878	0.548
Tram	0.6186	0.7595	0.6818

Table 8.4: Precision, Recall and F1 score per transportation modes for the RF classifier distinguishing between the four public transportation modes. The classifier is trained and tested on the MOASIS dataset using accelerometer features.

Mode (using accelerometer data)	Precision	Recall	F1
Bus	0.7903	0.8909	0.8376
Commuter Train	0.8125	0.7027	0.7536
Train	0.8462	0.8049	0.8250
Tram	0.8481	0.8481	0.8481

Appendix 2: Code

Table 8.5: Overview over scripts containing the relevant steps for the transportation mode classification.

Function	Description
GapFinding	Determines the time gaps between two recorded points in the trajectory. In determining a threshold, gaps that are too large can be detected. “newDiffTime” contains the information about the time difference between two entries. $findGaps(gps, timeCol = "newDiffTime", gapSize = 300)$
Change-point based segmentation	The change-point based segmentation is applied in creating a function with different statistical characteristics (getStat). Using the rollApply function these features are calculated over a certain window. In looking at the results of the values, the status column is changed. Based on the entries in the status column, the trajectory is segmented. $rollApply(GPS\_split[[i]]\$speed, getStat, window = 30, minimum = 15, align =' left')$
Transition-oriented segmentation	Determine the breaks along the trajectory. With the segment() function the transition points in a trajectory are determined, which are then used to split the trajectory into the segments. $for(iin1 : length(GPS\_split))\{Segments\_GPS[[i]] < -segmentation(GPS\_split[[i]], Kmax = 25, lmin = 30,$ $seg.var = "speed", scale.variable = FALSE)\}$
Fixed time window segmentation	The features are grouped by minute, and each minute is determined as one segment. The mode occurring most often in that minute is assigned to the entire time interval. $for(iin1 : length(all\_GPS\_list))\{$ $all\_GPS\_list[[i]] < -all\_GPS\_list[[i]]\% > \%$ $group\_by(byMinute = cut(as.POSIXct(timeConverted.x), "1min"))\}$
Feature selection	For each segment that has been determined in the previous step, the features are calculated in one loop using the following example code for each feature to be computed. Eventually, one large data frame containing the features of all segments and the mode is computed, which can be used for classification $as.data.frame(sapply(split(all\_BOTH\_list[[i]]\$speed, all\_BOTH\_list[[i]]\$segment), mean))$
Classification	The data frame containing the features is partitioned into the training and testing datasets (70% training, 30% testing). Using the 10-fold cross-validation and up-sampling, all the classifiers are trained. Using predict() the testing dataset is predicted by the classifiers and can be seen in the confusion matrix. $TrainingParameters < -trainControl(method = "repeatedcv", number = 10, repeats = 10, sampling =' up')$ $svm < -train(mode \sim ., data = trainingData\_GPS[, c(2 : 22)],$ $trControl = TrainingParameters, method = "svmRadial",$ $preProcess = c("center", "scale"),$ $tuneGrid = tuneGrid, metric = "Accuracy")$ $predict(svm, testingData\_GPS)$ $confusionMatrix(reference = testingData\_GPS\$mode, data = predicted, mode = "everything")$ For RF: $method = "ranger"$ For NN: $method = "monmlp"$

Appendix 3: Data structure

List of the most important columns in the files containing the measured information about the trajectories.

Table 8.6: Structure of the data used for the analysis.

Benchmark	Gpstime, time	Epoch timestamp
	dfname	Name of the trajectory
	Latitude	Latitude of GPS
	Longitude	Longitude of GPS
	Altitude	Altitude of GPS
	Speed	Speed of GPS
	Hdop	Horizontal Dilution of Precision (Quality of GPS signal)
	Mode	One of the seven transportation modes
	Note	Script-number of the trajectory
	Sensor	Sensor type of IMU
	X	Acceleration in x-axis
	Y	Acceleration in y-axis
	Z	Acceleration in z-axis
	Total	Total acceleration
	Time, Date	Time, date from Unix timestamp
MOASIS	Acc_X.mg	Acceleration in x-axis
	Acc_Y.mg	Acceleration in y-axis
	Acc_Z.mg	Acceleration in z-axis
	Acc_Total	Total acceleration
	Mode	One of the seven transportation modes
	Longitude	Latitude of GPS
	Latitude	Longitude of GPS
	Speed	Altitude of GPS
	ts.y	Unix timestamp

Personal declaration:

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in this thesis.

Zurich, 30th of January 2020
Location, Date

F. Corti

Federica Corti