Kick-scooters identification in the context of transportation mode detection using inertial sensors: Methods and accuracy

Fadoua Taia Alaoui, Hassen Fourati, Alain Kibangou, Bogdan Robu, Nicolas Vuillerme

RESEARCH PAPER

Kick-Scooters Identification in the Context of Transportation Mode Detection using Inertial Sensors: Methods and Accuracy

F. T. Alaoui[a] and H. Fourati[b] and A. Kibangou[b] and B. Robu[b] and N. Vuillerme[c]

[a]Univ. Grenoble Alpes, CNRS, GRICAD, F-38000, France; [b]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, F-38000, France; [c]Univ. Grenoble Alpes, AGEIS, F-38000, France

ARTICLE HISTORY
Compiled November 14, 2022

CONTACT H. Fourati. Email: hassen.fourati@gipsa-lab.fr

ABSTRACT
This work presents a novel transportation mode detection algorithm that handles the recognition of kick-scooters. In 2015, 10 minutes of data from a kick-scooter were considered in a transportation mode detection study, yielding a 56% F1-score. Since then, kick-scooters were not given much attention. Yet, kick-scooters are now very present in the urban transportation ecosystem, and their consideration in transportation studies has become a must. To fill this gap, 4 hours of kick-scooter signals were collected by 18 participants, with a set of 6 different kick-scooters, using 3 body-worn inertial measurement units. Obviously, kick-scooter patterns are classified in contrast with other modes of transportation. Two classification scenarios are considered in order to gradually increase the classification model complexity. The first scenario includes walking, biking, and kick-scooter, while the second considers public transport (tramway and bus) in addition to the former transportation modes. Results show that kick-scooters can be detected with an F1-score of 80% in the first scenario. Walking and public transport samples were still accurately classified in the second scenario, with an F1-score above 80% for both classes. However, bike and kick-scooter samples were both classified with lower F1-scores, equal to 59% and 64% respectively. Therefore, the main focus of future works should be directed towards the separability of kick-scooters and bikes when public transport is considered. The findings also suggest to place preferably the sensors in the trouser's pocket, allowing for leg motion to be finely captured.

KEYWORDS
Transportation mode detection/classification/identification; convolutional neural network; kick-scooters; micro-mobility; classification; inertial sensors

1. Introduction

1.1. The growth of micromobility

In the last few years, light-weight vehicles have become popular among adults and young people. Bikes B. Wang, Vu, Kim, and Cai (2022), kick-scooters (KS), monowheels, and segways have nowadays become part of a larger transportation landscape designated as micro-mobility systems. KS have shown several advantages over bikes and other micro-vehicles. For example, KS are generally smaller, lighter, and can be folded and carried in tramways, buses, and metros. Therefore, they constitute relevant candidates to solve the first mile last mile (FMLM) issue, i.e. the distance that separates home or the place of work from the nearest public transport (PT) station, see McKenzie (2020). In the same vein, the findings of Zuo, Wei, Chen, and Zhang (2020) suggest that transit accessibility can be significantly increased by the use of bikes over walking. In fact, the faster the reachability of a station, the more it is likely to be accessed. The same conclusion was made in Shiv (2018) where the lack of speedy mobility was found to be a catalyst for the use of private automobile to travel the FMLM distance. On the other hand, KS are green and health-friendly vehicles that are rather easy to handle. The PriestmanGood's kick-scooter for life [1] is a typical example illustrating the ease of usability of KS by the elderly, which in other respects allows them to safely maintain physical activity. Furthermore, KS are allowed to share bike lanes and other public spaces, see Kostrzewska and Macikowski (2017). Thus, their integration to urban transportation systems is effortless in terms of infrastructure, see Oeschger, Carroll, and Caulfield (2020). As a result of the worldwide success of KS, it appears to us that their inclusion in Transportation Mode Detection (TMD) solutions is a requirement. Not only this will enhance urban navigation tools, but it will also improve urban policies and shared micro-mobility systems, see Ashqar, Elhenawy, Rakha, Almannaa, and House (2022).

1.2. TMD families of methods

1.2.1. Sensors

Even though KS are new to TMD research, multiple TMD methods have been developed to recognize common transportation modes such as private cars, bikes, and PT. Recent TMD surveys can be found in Sadeghian, Håkansson, and Zhao (2021), Kamalian, Ferreira, and Jul (2022), Ahmed and Diaz (2022). Historically, the main devices used were first GPS loggers Biancat, Brighenti, and Brighenti (2014), Roy, Fuller, Nelson, and Kedron (2022) and then smartphones Carpineti, Lomonaco, Bedogni, Di Felice, and Bononi (2018), Sharma, Singh, Udmale, Singh, and Singh (2021), Wang, Luo, Zhao, and Qin (2021), Liu (2022), which are still in use at the present, although a recent study has explored the use of smartwatches to perform TMD, see Hasan, Irshaid, Alhomaidat, Lee, and Oh (2022). The chosen technologies used for TMD purposes, regardless of the device, have evolved from the use of standalone GPS receivers Zheng, Liu, Wang, and Xie (2008), often fused with Geographical Information Systems (GIS) Gong, Chen, Bialostozky, and Lawson (2012), Shah, Wan, Lu, and Nachman (2014), in favor of proprioceptive low-cost sensors due to their energy-efficiency and availability in off-the-shelf devices such as smartphones, see Yu, Yu, Wang, Lin, and Chang (2014). However, GPS receivers can still be used in favorable environments, typically in open-sky conditions, such as in Hadjidimitriou, Cantelmo, and Antoniou (2022), particularly when real-time functioning is not a requirement. Otherwise, the commonly used self-contained sensors are accelerometers, Nham, Siangliulue, and Yeung (2008), gyroscopes,

[1] https://www.priestmangoode.com/project/scooter-for-life/

Jahangiri and Rakha (2015), and magnetometers, Fang et al. (2016). Few works have used barometers Sankaran et al. (2014), microphones, L. Wang and Roggen (2019) and light sensors Carpineti et al. (2018). More exceptionally, bluetooth devices have been used in Coroamă, Türk, and Mattern (2019) to recognize transportation modes. Indoors, in a hospital-specific environment, Wireless Fidelity (WIFI) has also been used to classify hospital-specific modes of transportation, among which a KS, although the dataset comprised only 10 minutes of data from KS, see Prentow, Blunck, Kjærgaard, and Stisen (2015).

1.2.2. Time latency

On a methodological basis, there is a clear boundary in the literature between trip-based solutions as in Zhou et al. (2017), and near-real-time solutions, with a latency of recognition equal to the total duration of a trip in the former approaches, while it roughly varies between 1s to 10s in the latter, as in Wang et al. (2019), Prelipcean, Gidófalvi, and Susilo (2017). Overall, the choice has been made in favor of low-cost and energy efficient sensors, together with low-latency solutions, lasting over 5s in average Wang et al. (2019), in order to meet the requirements of real-time applications such as navigation apps.

1.2.3. Classification models

Part of this work is model selection. Among the algorithms used in the literature, Random Forest (RF) is one of the most popular machine learning methods used in TMD problems, see Fang et al. (2016); Carpineti et al. (2018); Lorintiu and Vassilev (2016); L. Wang, Gjoreskia, Murao, Okita, and Roggen (2018). More recently, deep learning methods were used for the same purposes, for example feed-forward artificial neural networks (ANN), see Fang, Fei, Xu, and Tsao (2017), convolutional neural networks (CNN), see Liang and Wang (2017), and long-short term memory (LSTM) neural networks, see Asci and Guvensan (2019). These models are generally compared, especially in recent studies such as Zarei Yazd, Taheri Sarteshnizi, Samimi, and Sarvi (2022). Let us recall though that the model selection here is performed based on two particular criteria. First, the chosen models are the most successful in the literature. Second, it is now well-established that the model selection is best realized based on cross-validation, see He, Zhao, and Chu (2021), and Zimmer, Lindauer, and Hutter (2021), which is also the approach followed in this study. Hence, RF, ANN, LSTM, and CNN were all tested systematically according to a 5-fold cross-validation process based on 1 separate and randomly selected test subjects for each fold. This way, the evaluation scenario adopted in this work guarantees enough robustness and is enhanced with respect to the major part of past TMD studies. In the latter, the train and test samples used to contain data from all subjects, see Taia Alaoui, Fourati, Kibangou, Robu, and Vuillerme (2022), resulting in an over-rated classification accuracy.

1.3. TMD datasets

As for publicly available datasets, sorted from most recent to oldest, the main datasets are the TMD-CAPTIMOVE, see Taia Alaoui et al. (2020), which is the dataset used in this study, collected by 34 participants and having an amount of 48 hours of data, the SHL dataset, Wang et al. (2019), collected by 3 participants with a total duration of 703 hours, the US-TMD dataset, see Carpineti et al. (2018), collected by 13 participants with a total duration of 32 hours, the dataset of Ballı and Sağbaş (2018), collected by 4 subjects with a total duration of 8 hours, and finally the oldest (2014), and at the mean time the richest in terms of sample size and amount of data, which is the HTC dataset, see Yu et al. (2014), collected by 224 participants with a duration of

8311 hours. With respect to the other publicly available datasets, the current dataset (i.e. TMD-CAPTIMOVE) is optimized in terms of balance between the number of participants and the total duration, as shown in Figure 1. The x-axis provides the total duration of the considered TMD dataset in a log10 basis, the y-axis provides the number of subjects, also in a log10 basis, while the size of the circles provides the ratio between sample size (i.e. the number of participants) and the total duration, also in log10 for a better readability and visualization. Through this ratio, we wanted to recall the importance of optimizing the balance between the sample size and the total duration, as cross-subject variability, Taia Alaoui et al. (2022), as well as time-dependency inside the data, Prentow et al. (2015), were found to be critical in TMD training problems.

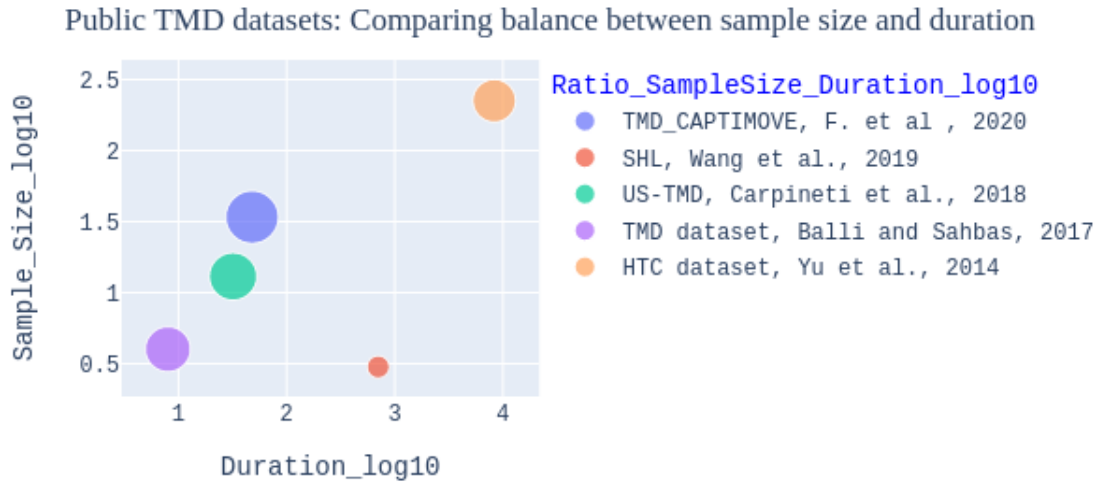


Figure 1.: Public TMD datasets: comparison of the balance between the number of participants and total duration

In Figure 1, the most optimal dataset is the one on the top right corner and with the largest circle size, and the least optimal is the one at the bottom left corner and with the smallest circle size. Thus, the TMD-CAPTIMOVE is the most optimal with respect to circle size, which is the balance between the number of participants and the total duration. At the mean time, it is undeniably less optimal than the HTC dataset in terms of duration and sample size, and so it is as compared with the SHL dataset in terms of duration. More globally, it is an improved TMD dataset as compared with the other remaining public datasets. It is also more optimal and more balanced than many other private datasets, which have lower sample size and duration, Zhao, Hou, Alrobassy, and Zeng (2019), with 11 participants and 15 hours of data, Delli Priscoli, Giuseppi, and Lisi (2020), with 18 subjects and 104 hours of data, and Su, Yao, He, Lu, and Tong (2017), with 12 subjects and 8 hours of data.

On the other hand, it is the first dataset that provides hours of data (between 2 and 4 depending on the class) from KS and from elevators and stair-cases. Thus, it considers a seamless TMD transition from the indoor to the outdoor and vice-versa. Besides, the time amount of collected data from these modes of transportation is naturally lower than the duration of other TMD classes such as PT or walking. In fact, the latter last longer on a daily basis. Therefore, the fact of systematically balancing the amount of data for each class, Liang and Wang (2017), most of the time at the cost of losing information, see Carpineti et al. (2018), seems questionable.

1.4. KS as a novel class in TMD

In parallel to the technological and methodological challenges tackled by TMD research, the focus of this study is the augmentation of the types of transportation modes being classified. This augmentation is directly obained through the consideration of KS among other common transportation modes. Indeed, the most popular TMD classes are tramway, see Byon, Ha, Cho, Kim, and Yeun (2017), bus, see Liang, Zhang, Wang, and Xu (2020), Catalán, Lobel, and Herrera (2022) and metro (or subways), see Fang et al. (2017). Bikes and e-bikes were considered in Xiao, Juan, and Zhang (2015), whereas one study, Prentow et al. (2015), has considered 10 minutes of KS data, in a specific hospital environment, resulting into a KS-F1-Score of 56%. The latter was the first contributing study to KS integration to TMD methods. It is granted for having explicitly emphasized the significance of KS in TMD and for having provided precise information on their dataset collection scenarios and results.

More globally, due to the lack of studies on KS-dedicated TMD, this work will not make a thorough revision of the pros and cons of past TMD approaches. Instead, we investigate the possibility to extend existing TMD approaches to the use of KS. Therefore, the main issues addressed in this research are questioning the ability for KS to be accurately detected among other transportation modes. Another important question is whether the integration of KS impacts the accuracy at which other transportation modes are detected. These questions are answered based on experimental results obtained according to two scenarios. The first scenario consists in the identification of KS as a separate transportation mode from 2 other locomotion modes: on-foot and bike. The second scenario, which is more complex, adds to the first one signals from bus and tramway, merged together and considered as one class called PT.

1.5. Contributions

The main contributions of this work can be summarized as follows:

- For the first time, KS signal patterns are thoroughly studied. Two major KS modalities could be distinguished: pushing and cruising. This shows that the training of KS is more challenging as compared with mono-modal transportation modes such as walking or driving a car.
- The placement of wearable sensors on the human body is studied as part of the variables to be optimized. Especially for KS, it is shown that the placement of sensors should not be random, which contrasts with past studies where TMD models were built regardless of the placement of the sensors on the body, see Carpineti et al. (2018).
- We demonstrate that detecting KS with an F1-score above 80% is possible when considering on-foot activity and biking. However, the accuracy at which KS is detected decreases significantly when PT is considered. It is shown that on-foot and PT are not negatively impacted by the inclusion of KS. In contrast, bike and KS patterns become more challenging once the classification contains PT data.
- The model validation is performed on data drawn from 5 randomly chosen test subjects (U1-U5), thus allowing us to evaluate the generalization ability of the model. Note that this aspect has been generally omitted in the TMD literature, see Taia Alaoui et al. (2022).

2. Materials and Methods

2.1. Data collection

The data were recorded by 3 Inertial Measurements Units (IMUs) that include a tri-axis accelerometer, a tri-axis gyroscope and a barometer each. Their respective placements were on the foot, on the waist and in the trousers' pocket. They can be visualized in Figure 2 in orange. Each participant was equipped with a front camera for verification purposes, as shown in Figure. 3. The data used in this study were sub-sampled at a frequency of 32Hz.

(a) (b) (c)

Figure 2.: Position of the wearable sensors on the body: Foot-mounted (a), in the pocket (b), waist-attached (c)

Figure 3.: Body-worn camera during data collection

The data were collected by 34 participants from young to older ages, 20 men and 14 women aged from 18 to 50 years old, who volunteered to perform the experiments. They came from different social and professional backgrounds and were healthy subjects. There was no specific selection of participants based on either social, professional, or health factors. Such selection approach for the human subjects may affect the level of bias that potentially underpins the sample. The data collection was realized under different weather conditions from November 2019 to January 2020 during workdays at different day times, including rush hours (8-9am, 12-2pm, 5-6pm). According to our recruitment procedure, it is undeniable that the models built to perform the classification of TMD in this study are expected to be efficient basically for signals collected during workdays and rather by healthy and active subjects. Therefore, new data should be collected to adjust the algorithm for other populations, for example the elderly and individuals with varying levels of capability, as gait patterns are subject to variation in those cases, see Osoba, Rao, Agrawal, and Lalwani (2019).

In order to perform the experiments, the participants were equipped in the office. Then, they were accompanied during the whole experiment by the person in charge of later labelling the data. For tramways and buses, the itinerary was fixed at the beginning. For walking, biking, and KS riding, the volunteers were free to choose their itinerary. The experiments ended again in the office where the participants made a small report about the trip they made.

From 18 subjects, we were able to collect data using KS. In this scenario, they were asked to switch legs while riding the KS in order to balance pushing and resting phases for each foot. In total, 6 different KS were used. Each participant used one of the 6 KS to collect data. The other transportation modes involved at least 20 different participants and included walking, biking, and PT (tramway and bus). The time duration relative to each transportation mode is given in Figure 4. As usual in classification problems with real data, the classes are not equally distributed. The choice in this study was to take this imbalance into account at the validation stage by providing the F1-Score as an accuracy metric instead of only providing the overall accuracy. Another choice could be to balance classes by removing data, which the authors of this study think is not optimal as this inevitably results in information loss.

The dataset can be accessed at https://perscido.univ-grenoble-alpes.fr/datasets/DS310.

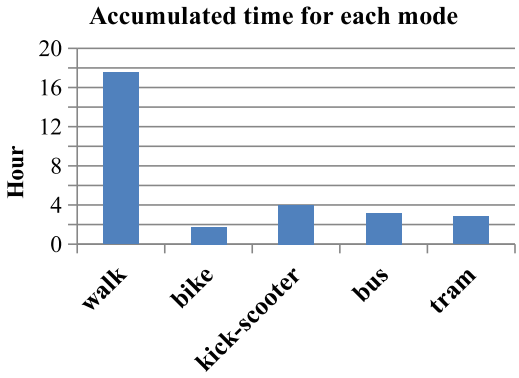


Figure 4.: Time duration relative to each transportation mode

2.2. Classified transportation modes

We adopt two different classifications of the transportation modes. They are given in Figures 5 and 6. The first classification consists of 3 activity states that are: walking (W), biking (B), and KS riding. This scenario corresponds to body-induced motion. It is more suitable for short distance trips and for people who use their scooter or bike as a principal means of transport. The second classification integrates PT, which is the result of merging tramway and bus data. This scenario has more to do with the FMLM distance as KS are usually used to reach PT stations.

(a) (b) (c)

Figure 5.: First set of motion modes. (a): walk (b): bike (c): KS

2.3. Signal visualization and analysis

Figures 7, 8, 9, 10, and 11 are examples of the signal patterns relative to each transportation mode. Note that the signals for the three sensor positions are not exactly synchronized and should be considered independently from one another.

(a) (b) (c) (d) (e)

Figure 6.: Second set of motion modes. (a): walk (b): bike (c): KS (d): tramway (e): bus

Over all transportation modes, a systematic difference distinguishes the foot-mounted sensor from the waist-attached sensor and the one in the trouser's pocket. The signal magnitude is much higher in the first case. This difference is however less important for PT as the subjects generally stay still inside vehicles, which supposes that their feet are still as well. A slight magnitude difference is though observed between buses and tramways due to higher vibrations induced by the road topography as compared with rails for tramways (Figure 10). In this work, both transportation modes (tram and bus) are considered in the same class, PT.

For biking (Figure 8), the acceleration and angular rate signals show a clear time-periodicity that reflects cycling. Yet, this cyclic pattern tends to fade in case the subject cruises for a period of time, for example between 4s and 7s.

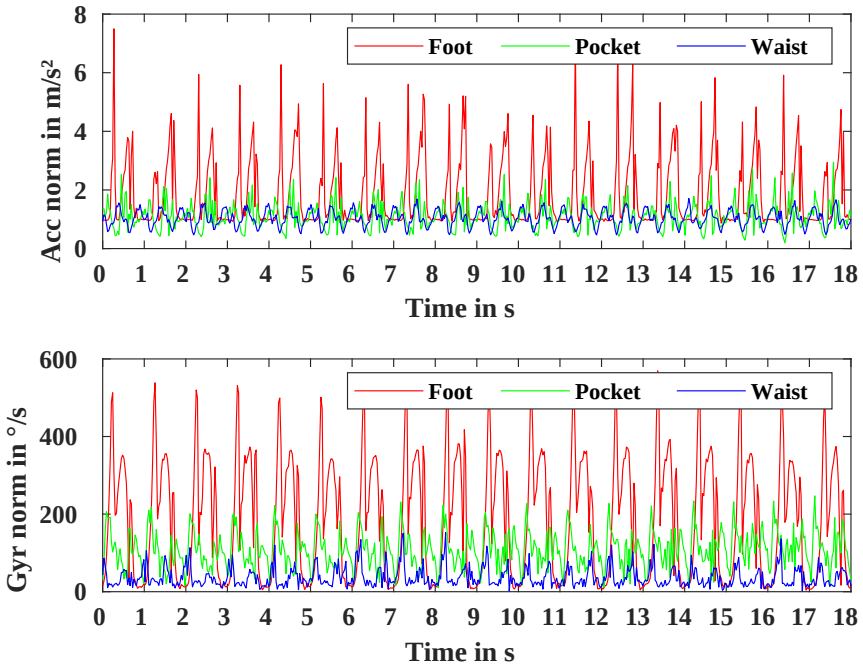


Figure 7.: Walking signal patterns: Acceleration norm in m/s^2 and angular rate norm in °/s

For KS, Figure 11 shows a signal frame that consists in two different KS patterns. The first pattern corresponds to pushing the ground with the foot on which the sensor is mounted. This motion is reflected by two consecutive signal peaks in both acceleration and angular rate. In fact, the first peak is the motion forward that allows the subject to gain momentum, while the second is the backward motion of the pushing foot before it hits the ground. The second pattern corresponds to cruising with both feet resting on the KS deck. These phases show flatter signals with small variation induced by ground friction while the scooter keeps moving by inertia.

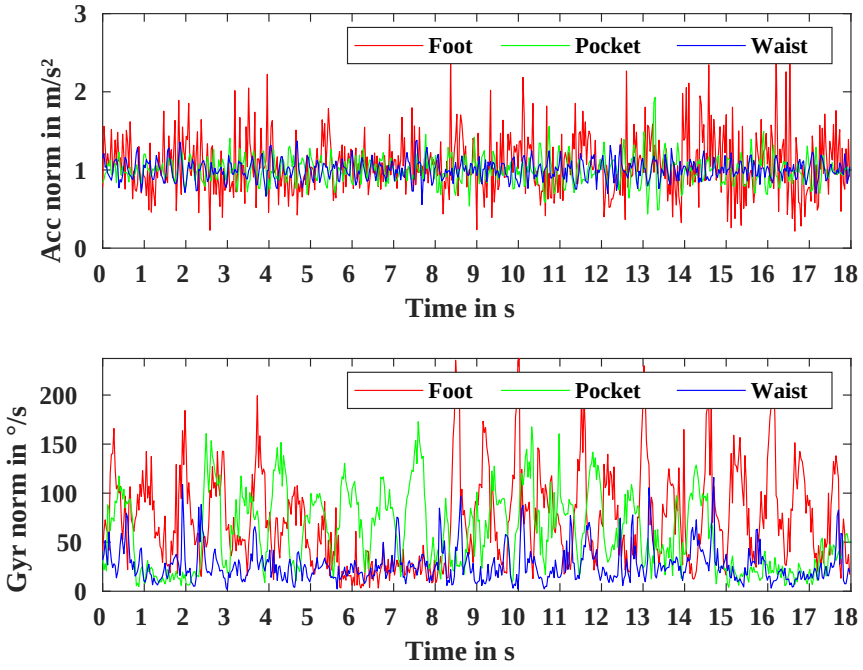


Figure 8.: Biking signal patterns: Acceleration norm in m/s^2 and angular rate norm in °/s

2.4. Signal pre-processings and segmentation

Signals were preprocessed prior to any segmentation or feature computation. The whole process of signal segmentation and preprocessing was assisted by videos that were time-stamped and synchronized with the IMUs. The participants were asked to behave freely during experiments. For example, they could move inside PT, walk in-between scooter riding phases, etc. In order to take this freedom into consideration, a peak detection and deletion process was undertaken in order to remove sporadic motions that are not related to the transportation modes. For both accelerometer and gyroscope 3D components, Tukey's schematic boxplots, see Dawson (2011), were calculated for multiple 1s long signal frames in order to detect outliers. The latter were then replaced by the upper and lower limits of the boxplot. Afterwards, the preprocessed 3D signals were used to compute the norm of acceleration and angular rate, which solves the problem of sensor orientation while later training the classifier.

In order to segment signals, we sought to minimize the segmenting window length while keeping satisfying classification accuracy, so that the detection time latency is minimized. To this purpose, many window length values, including 1s, 1.5s, 2s and 3s, were used to train 4 classifiers and validate them on a random test subject. The optimal window length value is that for which the classification accuracy is maximal. According to this approach, in the first classification scenario, this length was fixed to 1.5s. Note that it is much lower than the average sliding window length found in the state-of-the-art, which is about 5s, see Wang et al. (2019). For the second classification scenario, this value was set to 3s.

2.5. Classification Methods

2.5.1. Input data structure

In the training design, there are M train instances or train samples. In other words, M is the number of labeled signal patterns used for training. P is the number of signal

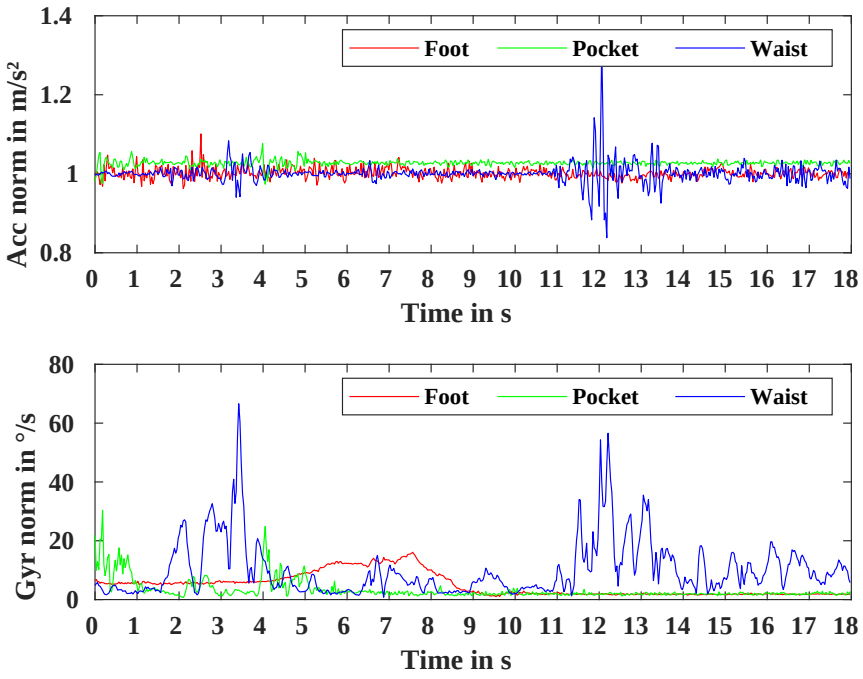


Figure 9.: Tramway signal patterns: Acceleration norm in m/s^2 and angular rate norm in °/s

samples in one train instance, equal to the product of the segmenting window length and the signal sampling frequency (see Eq 1). Q is the number of signal channels considered for training. For example, if acceleration and angular rate signals are used, then Q is 2. Q is also the depth of the 3D matrix used to train the CNN and LSTM models described later in this section. S is the number of computed features based on the relevant signal channels (see section 2.6.2). The number of classes is N.

$$P = Segmenting\,Window\,Length * Signal\,Sampling\,Frequency \tag{1}$$

In the following sections, the classification algorithms are provided in details. Note that there are two separate approaches used depending on the classification algorithm. For RF and ANN, the input data are 2D matrices of size MxS. In fact, these algorithms do not take into consideration the structure inside each train sample. Hence, computing descriptive features to train these 2 models results in a lower computational cost and higher training speed. For CNN and LSTM, the algorithms are intrinsically designed for structured data sequences. Therefore, no feature computation is performed with these 2 models. The input data for CNN and LSTM are time sequences, which means 1D vectors of length P. Yet, as Q signal channels are considered, an input train sample is a Q-dimensional vector of length P. Thus, the global input data of the LSTM and CNN models is a 3D matrix of size $M * P * Q$.

2.5.2. Random Forest

Random Forest (RF) is an ensemble of decision trees. The final prediction of a RF algorithm is computed through a majority voting based on the prediction of each decision tree. RF is known to prevent overfitting thanks to the bootstrap aggregation strategy that consists in a random selection of features and train samples to build each decision tree. The algorithm used in this study has 30 trees built upon a subset of 15 features each. The split criterion is the Gini index. The RF model is shown in Appendix A, in Figure 1.

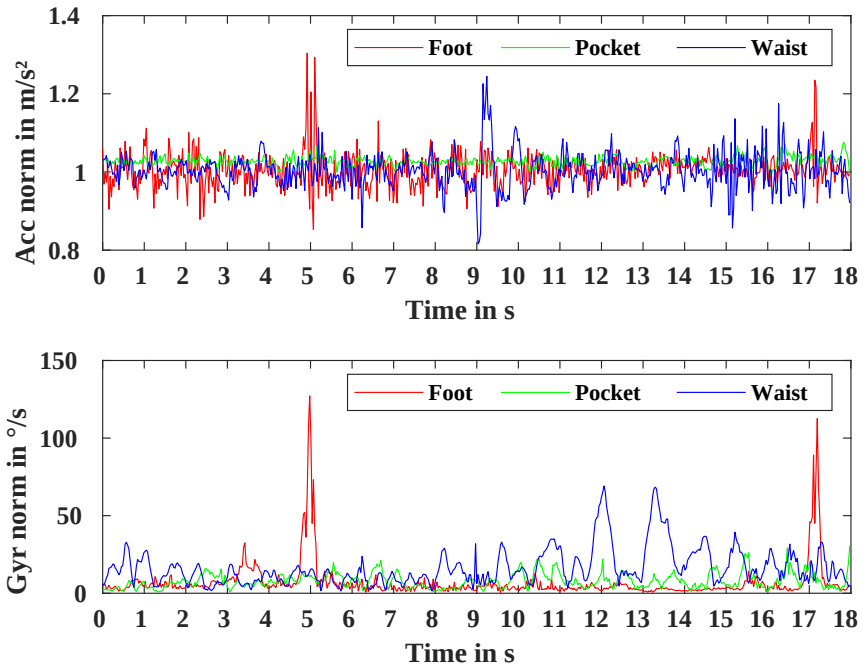


Figure 10.: Bus signal patterns: Acceleration norm in m/s^2 and angular rate norm in °/s

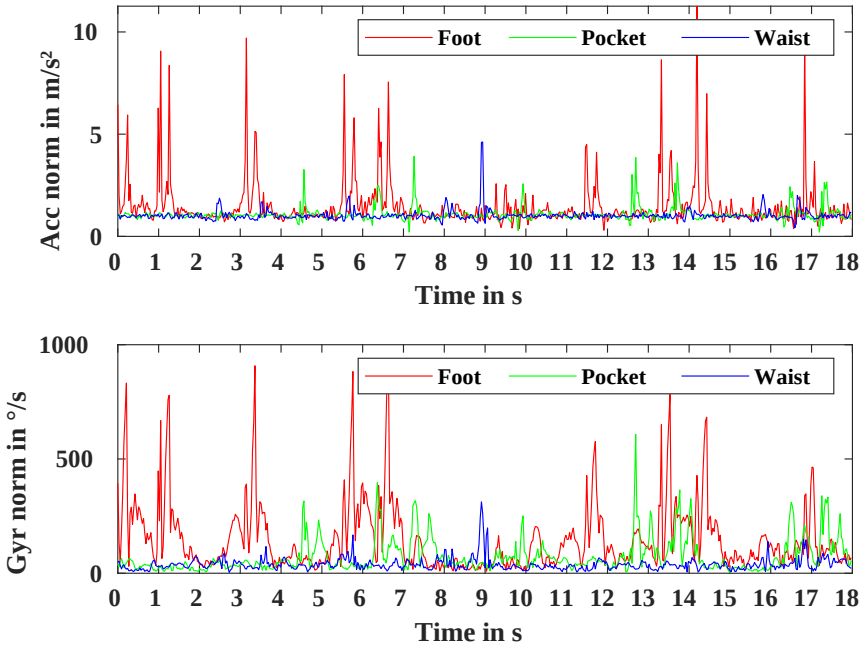


Figure 11.: KS signal patterns: Acceleration norm in m/s^2 and angular rate norm in °/s

2.5.3. Feed-forward Artificial Neural Network

The Feed-forward Artificial Neural Network (ANN) used in this study has two hidden layers of 50 and 100 neurons respectively. Naturally, the output layer has as many neurons as trained classes, which is N. The activation function of the input and hidden layers is Relu. The output layer operates according to a softmax activation function. Layer 2 and the output layer are optimized under an L2 regularisation with a lambda coefficient of 0.001. Each layer of the network is followed by a batch normalization

operation to stabilize the parameters distribution through the training process. The optimizer is Adam with an initial learning rate of 0.01 and the cost function is the categorical cross-entropy. The ANN network architecture is shown in Appendix A, in Figure 2.

2.5.4. Long Short Term Memory cell

Long Short Term Memory cells (LSTM) are designed for time series in order to learn memory and sequential structure of data. The LSTM model used in this study has 1 LSTM cell, a hidden layer of 100 neurons under a Relu activation function and L2 regularizer, and finally, an output layer of N neurons under a softmax activation function. The loss function is the cross-entropy, and the optimizer is Adam with an initial learning rate of 0.001. The model architecture is shown in Figure 1 of Appendix B.

2.5.5. Convolutional Neural Network

The Convolutional Neural Network (CNN) is composed of 3 convolutional layers of 30 filters (kernels) each. The kernels size for each layer are respectively P/8, P/4, and P/4. A Maximum pooling of 2 was realized to the output of each convolutional layer. Two fully connected dense layers of 100 and 50 neurons follow the convolutional layers. The convolutional layers activation function was LeakyRelu with an Alpha value of 0.1. The activation function of the dense layers was ReLu. The output layer has N neurons and was designed with a softmax activation function. The cost function was the categorical cross-entropy. The optimizer was Adam, with an initial learning rate of 0.001. The global architecture of the CNN is shown in Figure 2 of Appendix B.

2.5.6. Computational tools and Transferability

RF was implemented based on Scikit learn library in Python. ANN, CNN, and LSTM were implemented using Keras library in Python as well. All computations were realized on an Intel Quad-Core i7 CPU of a laptop with a 16 Gb RAM memory. The operating system was Windows 10. All used libraries are free and can also operate under Linux and MacOS operating systems. As all the hyper-parameter values are provided, together with the models architecture, set of input features, and the online dataset. Therefore, the study is entirely reproducible.

2.6. Test scenarios and evaluation method

2.6.1. Test scenarios

Two main aspects are evaluated. First, we want to select the best suited features and signal inputs. Second, we want to select the most optimal model given the optimal input set. It is chosen among the selected algorithms from the literature: RF, ANN, LSTM, and CNN. For both purposes, cross-validation is used. It consists in training the models using 70% of the labeled samples, and then testing the stabilized models, i.e. after the training is done, to predict the labels (or classes) of the remaining 30% of the labeled samples. As commonly adopted in training problems, we performed a 5-fold cross-validation process. Indeed, the multiple-fold cross-validation process allows the evaluation of model stability against small variations in the training set. In this study, each fold contains data from all subjects except one, randomly chosen and labeled from U1 to U5. It is used to validate the model.

A confusion matrix is provided at the end of each cross-validation. It provides the prediction results for all samples from the test set. This enables the computation of

prediction errors and the computation of numerous accuracy metrics. Appendix C is provided with all confusion matrices. Details on the accuracy metrics used in this study can be found in section 2.6.4. Once these metrics are calculated, the choice has been made in this work to keep the model that leads to the best classification accuracy of KS.

Model selection is realized according to a two-stage scheme. First, we have trained the models using separate inputs, that is the selection of relevant input signal channels. Second, we have discarded the irrelevant signal channels, and trained the four algorithms using the optimal feature set, which allowed us to select the best model architecture. As the input selection phase is an intermediate step towards the final model selection, the methodology used to perform this selection is described in the section 2.6.2. Thus, the results in section 3 only present the performance of the four classification models used with the optimal set of inputs.

2.6.2. Selection of the optimal input set

Model selection requires first to select the relevant data to be used for training. Here, multiple signal channels were considered. They are provided in column 1 of Table 1. While most of the computed features are quite common, the energy might be questioned. It is defined in the time and frequency domains by equations 2 and 3 respectively.

$$Energy([t1, t2]) = \int_{t1}^{t2} s(t)^2 dt \tag{2}$$

where t refers to time, and s to the considered signal. t1 is the beginning of the time interval while t2 is its end.

In the frequency domain, the same formula is used, using frequencies instead of time intervals:

$$Energy([f1, f2]) = \int_{f1}^{f2} s(f)^2 df \tag{3}$$

where s is now the signal expressed in the frequency domain. f1 and f2 are respectively the minimum and maximum limits of the considered frequency interval.

For each signal channel, a classification was realized using LSTM and CNN algorithms. The signal channels with the lowest accuracy were removed while the others were taken together. The new set of signal channels was used with both algorithms. Then, each signal channel was iteratively removed. In case this led to a decrease in classification accuracy, the channel was fed back as an input to the neural network and another channel was removed to make a new test. In case removing the channel resulted in equal or higher classification accuracy, the channel was considered either useless or counter-productive, and therefore was discarded definitively. According to this analysis, the optimal set of signal channels was found to be composed of the gradients of acceleration and angular rate in both classification scenarios.

After the signal channels selection, descriptive time-domain and frequency-domain features were computed in order to be used with RF and ANN. They are provided in column 2 of Table 1. Note that the same set of descriptive features were computed for every channel.

Table 1.: Set of computed signal channels and descriptive features

Signal channels	Features computed for all channels
1) Acceleration norm, 2) Gradient of the norm of acceleration 3) Integral of the norm of acceleration 4) Angular rate norm 5) Gradient of the norm of angular rate 6) Integral of the norm of angular rate	Mean, Energy, Standard deviation, Variance, Main frequency component, Power of the main frequency component, Spectral energy between 0 and 2Hz, Spectral energy ratio between 0 and 2Hz, Spectral energy between 2 and 4Hz, Spectral energy ratio between 2 and 4Hz, Spectral energy between 4 and 10Hz, Spectral energy ratio between 4 and 10Hz, Spectral energy between 10 and 30Hz, Spectral energy ratio between 10 and 30Hz, Spectral centroid, Spectral spread, Entropy, Number of zero crossings

2.6.3. Selection of the best model

The optimal input set, given in the previous section, is used in the four classification models to be compared. This comparison is based on the chosen accuracy metric, which is the F1-score relative to KS. In other words, the best performing algorithm is the one that presents the highest accuracy for KS. For practical reasons, the confusion matrices, which contain all the details of the classification results, are provided only for the best model in Appendix C. However, the results in chapter 3 provide the global accuracy metrics for all classification models to visualize the performance obtained for each user. According to the test scenarios presented in section 2.6.1, the chosen model is the one that has the highest accuracy in average, taking into account the prediction results for 5 different test subjects (U1 to U5).

As a reminder, for each classification scenario, there are 3 confusion matrices in columns (foot, waist, pocket) $*$ 5 test subjects in row. For example, the first row of Appendix C consists of 3 confusion matrices obtained when testing the model on user 1 (U1), for the foot, the waist, and the trouser's pocket. The abbreviation for each class is: W: Walk, B: Bike, KS: Kick-scooter, PT: Public Transport. The statistics provided together with the confusion matrices are the OA (Eq. 5) as well as the geometric mean of the F1-score (Eq. 6). The imbalance of the dataset is taken into account by the F1-score metrics. The geometric mean of the F1-score provides information on the global accuracy at which each class has been learnt. In other words, a high geometric mean of the F1-score indicates that all the classes were accurately learnt. On the contrary, a weak geometric mean of the F1-score indicates that at least 1 class has been weakly learnt, giving it the same weight as the other classes in spite of their unequal time distribution.

2.6.4. Accuracy Metrics

The accuracy metrics are chosen based on the questions that this study addresses. The first is to know how precise is KS detection among different urban transportation modes. The second is to know whether the integration of KS in TMD models makes the recognition of other transportation modes less accurate. Therefore, the classification accuracy results are given relatively to KS. Indeed, the first question is answered by the F1-Score, Eq (4), relative to KS. In order to answer the second question, global accuracy metrics are also provided. They correspond to the overall accuracy (OA), Eq (7), and the geometric mean of the F1-Score, Eq (8), provided in Tables 1 and

2. The latter are global metrics that provide the classification accuracy of all classes taken together. In case the OA and geometric mean of the F1-Score are high enough (>70%), this means that all classes have been accurately detected. Otherwise, the F1-Scores of the classes other than KS are additionally analyzed in order to understand the errors of the classification model.

$$F1\text{-}Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

where the *Precision* is given by

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{5}$$

and the *Recall* is given by

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{6}$$

$$OA = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \tag{7}$$

$$Geometric\text{-}Mean(F1\text{-}Score) = \sqrt[N]{\prod_{k=1}^{N} F1\text{-}Score(k)} \tag{8}$$

3. Results

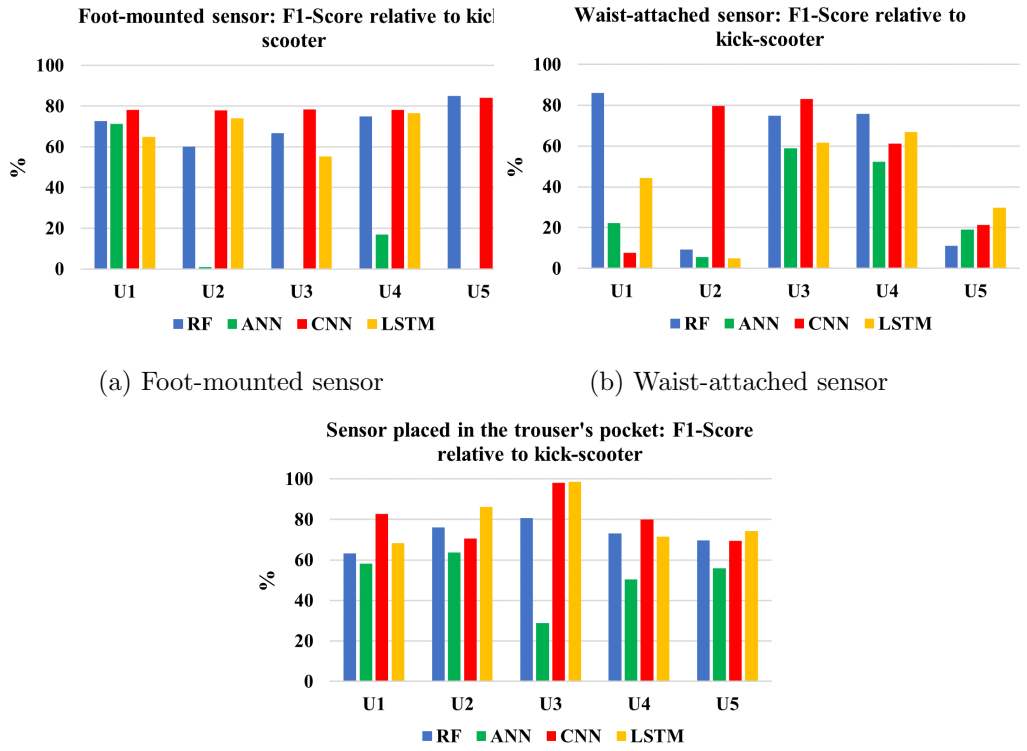
3.1. Model selection results

3.1.1. Model selection for scenario 1

In this section, several models are compared based on the F1-Score of KS for the 3 sensor placements. They are shown in Figure 12. The evaluation is made for 5 different test users (U1 to U5) that are not involved in training. From these figures, several conclusions can be drawn. They are summarized below:

(1) The highest F1-Score values are obtained using the CNN model with a foot-mounted sensor. In fact, the F1-Scores of all subjects are higher than 70%.
(2) The sensor placed in the pocket presents high F1-Score values for both the CNN (>70%) and LSTM (>65%) models. However, this performance varies significantly depending on the considered test subject.
(3) The F1-Scores relative to KS are the lowest for the waist-attached sensor. In fact, a high sparsity is observed over the different test subjects and different classification models.
(4) ANN has very weak performance over all subjects and all sensor placements. In fact, this does not necessarily mean that ANN models are less accurate than the other models. However, this result could be interpreted as a higher need for the tuning of the ANN hyper-parameters as compared with LSTM and CNN.

Given this analysis, it seems that the CNN model has the best performance in the first classification scenario.



(a) Foot-mounted sensor

(b) Waist-attached sensor

(c) Sensor placed in the trouser's pocket

Figure 12.: Model comparison for scenario 1

3.1.2. Model selection for scenario 2

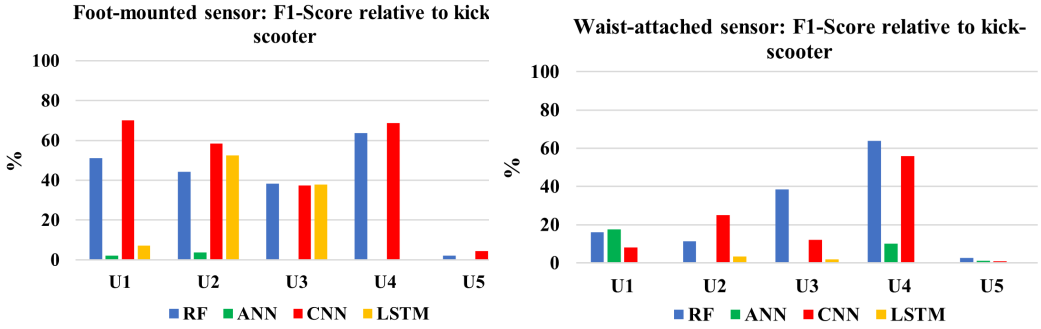
For the second scenario, the F1-Scores of KS are shown in Figure 13. Here, the performance is globally much lower as compared with the first classification scenario. One common observation to the 3 sensor placements is that U5 has very weak F1-Score for all classification models. Therefore, this result cannot be explained by the choice of the training algorithms but seems to be rather related to the data. A possible hypothesis is that U5 has atypical signal patterns (ex. patterns that depend on physiological features). Therefore, as the data from U5 were removed from the training set, the algorithm did not learn these patterns. Another reason could be that U5 has collected more data than the other test subjects. When looking at the confusion matrices, this hypothesis has been confirmed. In fact, U5 has contributed to the collection of PT samples much more importantly than the other users, i.e. U1 to U4. Therefore, U5 seems to be a crucial subject to the training and should not be removed from the training dataset. As a consequence, the classification results from U5 should not be considered in model selection.

Removing U5 from our evaluation, considering only subjects from U1 to U4, the CNN has again the highest average F1-Score as compared with the other models. The same optimal sensor placements as for the first classification scenario are found here, on-foot, and in the trouser's pocket.

3.2. Performance of the best classification model: CNN

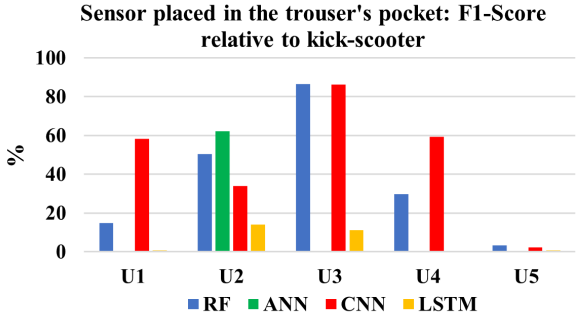
3.2.1. Scenario 1

After the selection of the relevant signal channels to be used for training, and after selecting the most performing algorithm in both classification scenarios, the present



(a) Foot-mounted sensor

(b) Waist-attached sensor



(c) Sensor placed in the trouser's pocket

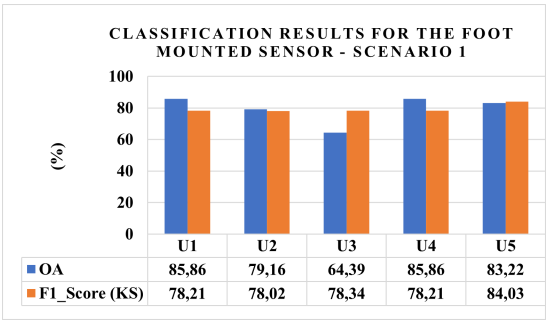
Figure 13.: Comparison of different classification algorithms in the second classification scenario

section provides more details about the achieved accuracies for KS specifically but also for the other transportation modes. To this purpose, Figure 14 shows the classification OA as well as the precise F1-Score values relative to KS for each sensor placement.

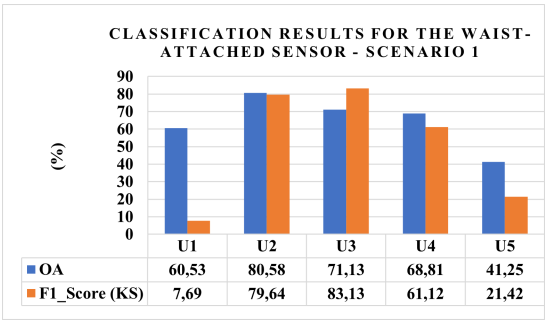
For the foot-mounted sensor, the F1-Score of KS varies from 78.02% to 84.03%. Its standard deviation is 2.61, meaning that the model is robust enough and can handle data from new subjects. The OA is comprised between 64.39% and 85.86%. In order to understand the lower OA value of 64.39% for U3, the confusion matrix for the foot-mounted sensor in Table 1 is analyzed. In fact, U3 has collected data only using a KS. Among these data, 2410 samples were classified as KS, 1324 were classified as biking, while 9 samples were classified as walking. In other words, the major errors here are due to confusing biking data samples with KS samples. More globally, the same tendency is observed for the other test subjects, however, it is most important for subject U3.

Surprisingly, even if the sensor placed in the trouser's pocket did not show strong signal patterns as compared with the foot-mounted one, it led to considerably high detection accuracy rates relative to KS, with an F1-Score comprised between 70.56% and 98.02% with a standard deviation of 11.52%. While U3 has an F1-Score relative to KS of 98.02% and an OA of 96.12%, the lowest OA is obtained for U5. Analyzing again the corresponding confusion matrix, it is observed that U5 has an OA of 60.64% and a geometric mean of the F1-Score of 61.39% for the sensor placed in the trouser's pocket. Looking more deeply into the confusion matrix, we can observe that around 1/3 of KS samples have been classified as biking.

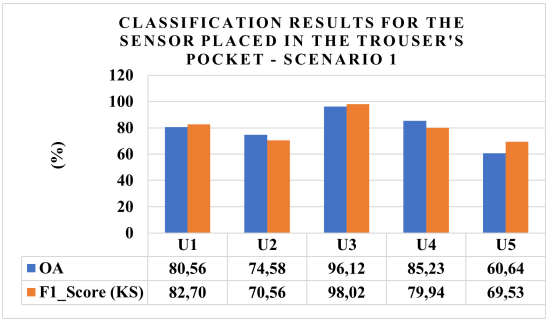
Finally, the F1-Scores obtained with the waist-attached sensor are comprised between 7.69% and 83.13%. Without any further analysis of these results, the waist seems obviously not to be an optimal place to recognize KS.



(a) Foot-mounted sensor, CNN



(b) Waist-attached sensor, CNN



(c) Sensor placed in the trouser's pocket, CNN

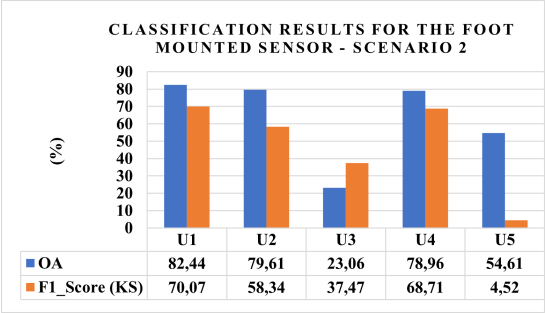
Figure 14.: CNN classification results for the first scenario

3.2.2. Scenario 2

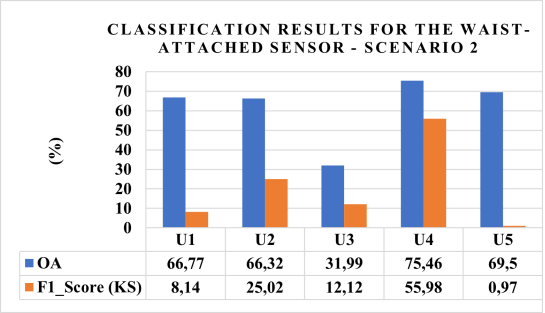
The results of the second scenario are shown in Figure 15. These results are given together with the detailed confusion matrices provided in Table 2. In this scenario, the global performance is lower with respect to the first scenario. As a reminder, U5 is not to be considered to interpret the classification results. However, the detailed statistics relative to U5 are still provided in order to understand the reasons why the accuracy is low for U5. More precisely, U5 made long travel periods of time by tramway and bus. Therefore, data from U5 were crucial to train the algorithm, which explains the low accuracy obtained when U5 is isolated as a test subject. This point can be clearly observed in Table 2 for U5.

Based on the results of the left test subjects, the F1-Score relative to KS ranges from 37.47% to 70.07% for the foot-mounted sensor, with a mean of 58.65% and a standard deviation of 15.6, which is an average performance that still requires post-classification improvements. For the sensor placed in the pocket, the F1-Score ranges from 33.86% to 86.24%. It has a mean of 59.35% and a standard deviation of 21.40%. The same conclusions as for the foot-mounted sensor are drawn. Finally, the waist-attached sensor presents weak F1-Score values with an average F1-Score of 25.32 and a standard deviation of 21.68%.

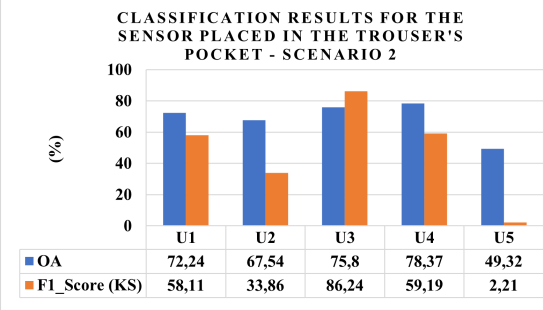
The study of the F1-Scores of KS for the second scenario shows that KS is detected with a weak accuracy when PT is included to the classification, namely tramways and buses. A deeper analysis of the errors in scenario 2 is realized in the discussion section.



(a) Foot-mounted sensor



(b) Waist-attached sensor



(c) Sensor placed in the trouser's pocket

Figure 15.: CNN classification results for the second scenario

4. Discussion

4.1. Classification model selection

In this work, 4 different algorithms were compared with the objective of providing a system that recognizes KS with enough accuracy. The tested algorithms were selected from the literature as mentioned in the introduction, with RF, ANN, CNN and LSTM. Selecting the relevant signal channels was crucial in this work. In this regard, the process of building several classifiers based only on one signal channel each has resulted in optimal performance. The validation of each algorithm was performed using a leave-one-subject out 5-fold cross-validation approach, meaning we have repeated the operation 5 times with 5 different test subjects. Therefore, the model validated in this study is robust enough to deal with data from unknown subjects. Note that we could isolate more subjects for testing, however, the dataset should be more significant in that case, especially as all subjects did not equally participate to data collection.

Based on these considerations, CNN was found to be the most optimal algorithm, with high classification accuracy rates and low standard deviations, which is an indicator of robustness.

4.2. Accuracy of KS detection

4.2.1. First classification scenario

The results show that it is possible to detect accurately KS out of walking and biking patterns using a sensor that is either foot-mounted or placed in the trouser's pocket, which is the answer to the first question of this study. For the foot-mounted sensor, the average F1-Score relative to KS is 79.36%, with a standard deviation of only 2.61%. For the sensor placed in the trouser's pocket, the F1-Score average relative to KS is 80.15% and its standard deviation is 11.52%. In this case, the increase in the F1-

Score is insignificant, while the standard deviation is 5 times higher than the standard deviation of the foot-mounted sensor.

In order to address the second question of this study, which is about the impact of KS on the detection of other transportation modes, the geometric means of the F1-Score given in Table 1 and noted "F1-Score" for simplicity, are analyzed. For the most optimal sensor placement, which is on-foot, the average geometric mean of the F1-Score (which is averaged over all 5 subjects U1 to U5) is 82.56% and its standard deviation is 3.37. In fact, this is a significantly high value that means all transportation modes are accurately classified regardless of their distribution. Hence, introducing KS with walking and biking does not result in a weakened detection accuracy of these classes. As a final conclusion, these results suggest that KS can be detected accurately and therefore distinguished from walking and biking patterns, while keeping a highly satisfying detection accuracy of the latter.

4.2.2. Second classification scenario

In the second classification scenario, as PT is included to the classification, the F1-Score relative to KS decreases sharply. The mean F1-Score relative to KS is 58.65% for the foot-mounted sensor, 25.32% for the waist-attached sensor, and 59.35% for the sensor placed in the trouser's pocket. In other words, the detection accuracy of KS in the second scenario is rather low as compared with the first scenario.

The geometric mean of the F1-Score for the foot is 68.28% with a standard deviation of 20.56, while the geometric mean of the F1-Score for the pocket is 70.35% with a standard deviation of 14.11. In fact, the global classification accuracy is much higher than the F1-Scores relative to KS. In other words, transportation modes other than KS have been satisfyingly classified. In order to confirm this suggestion, the F1-Scores of all considered classes are analyzed. They are shown in Tables 2, 3, and 4.

Table 2.: F1-Scores of each class of the second scenario for the foot

Foot	**KS**	**PT**	**W**	**B**
Average F1-Score	58.65	91.59	89.67	67.21
Std F1-Score	15.06	1.21	9.83	3.8

Table 3.: F1-Scores of each class of the second scenario for the waist

Waist	**KS**	**PT**	**W**	**B**
Average F1-Score	25.32	83.02	78.97	46.76
Std F1-Score	21.68	10.09	27.51	20.92

Table 4.: F1-Scores of each class of the second scenario for the trouser's pocket

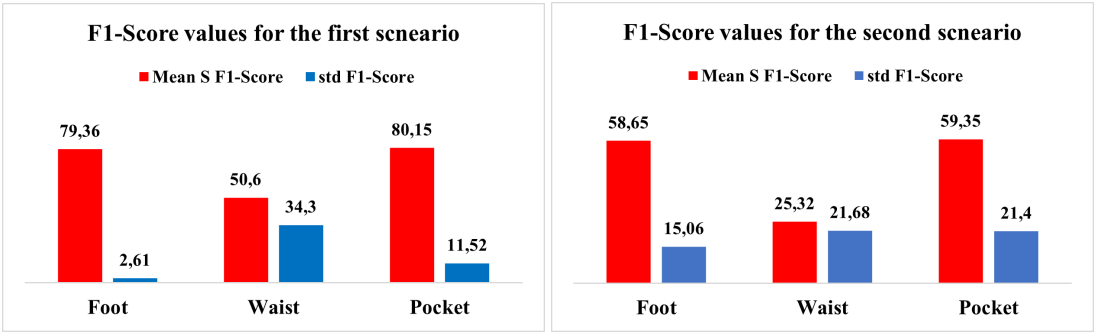
Pocket	**KS**	**PT**	**W**	**B**
Average F1-Score	59.35	85.26	81.4	64.37
Std F1-Score	21.4	7.55	15.93	11.24

From these tables, it is observed that PT samples were overall accurately classified,

whilst the main classification errors consisted in classifying KS samples as biking, or biking samples as PT. In both cases, this is about confusing weak activity phases for biking and scooter riding with being inside PT. In fact, these are phases where the subjects are more likely to be cruising with both feet resting on the KS deck or on the bicycle, which may be confused with being inside PT with a low-energy profile. In fact, in both cases, acceleration and angular rate magnitudes are weak. Besides, it is important to note that the data were labelled as bicycling and KS riding even when the participants were in a static mode on their bike or KS, for example while waiting at a street light, which enhances these confusions. As a global conclusion about the detection performance in the second scenario, the introduction of PT results in a decreased detection accuracy of both biking and KS. In this case, it is not the introduction of KS that has a negative impact on other classes, but rather the introduction of PT that makes the classification of bike and KS less accurate.

4.3. Optimal sensor placement

This section recapitulates the classification results in both scenarios and for all sensor placements and all test users. The mean and standard deviation of the F1-Scores relative to KS are provided in Figure 16. As this is obviously observed in the figure, the foot is the placement that offers the best trade-off between F1-Score value and standard deviation. Note that the latter is supposed to be minimized while the F1-Score is sought to be maximized. Therefore, even if the pocket has a higher mean F1-Score relative to KS, it has much higher standard deviation, which reflects much higher values for the F1-Score for some test subjects, but also much lower values, which is to be avoided for real life applications. As a result, a foot-mounted sensor is recommended in studies that seek to optimize the robustness of KS detection.



(a) Model comparison scenario 1 (b) Model comparison scenario 2

Figure 16.: Classification results for the 3 sensor placements

5. Conclusion

This work was a first contribution to the detection of KS among other urban transportation modes with a new improved and publicly released TMD dataset. Our first focus was about the choice of the relevant signal channels, and classification algorithm, which came out to be the gradient of acceleration and angular rate used with an Adam-gradient descent CNN algorithm. In the first classification scenario, we were able to classify KS samples with a mean F1-Score of 80.18% for the sensor located in the trouser's pocket and 79.36% for the foot-mounted sensor. Moreover, these values were computed based on a 5-fold subject-based cross-validation method, which takes into account the model reliability with regard to handling data from various test

subjects that were not involved in the training. Yet, once PT patterns were included, the detection accuracy decreased. In this case, an improving perspective could be to consider a pushing mode and a cruising mode for KS, as well as a cycling mode and a cruising mode for bikes, in order to decrease the intra-class variance, which is also a variable to be minimized while building robust classifiers. In summary, this study has been fulfilled in that it provided answers to the aforementioned questions. We proved that detecting KS is feasible, although it gets more challenging when PT is considered. Besides, results have shown that despite the difficulty to detect KS while including PT, PT were still accurately classified, meaning that KS inclusion in TMD does not result in a decrease of the detection accuracy of other transportation modes. An additional recommendation is to place the sensors in the trouser's pocket, or on the foot, in order to capture leg movements during KS riding. Finally, future contributions should also consider extending the sample distribution to include more varying signal patterns: healthy elderly, disabled, as well as signals collected during the weekends and holidays in order to study the time dependencies that may require necessary algorithmic adjustments.

Acknowledgement(s)

The graphical content used in this paper was designed by Freepik and Flaticon.com.

Disclosure statement

The authors declare no conflict of interest.
natbib

References

Ahmed, D. B., & Diaz, E. M. (2022). Survey of machine learning methods applied to urban mobility. *IEEE Access*, *10*, 30349–30366.

Asci, G., & Guvensan, M. A. (2019). A novel input set for lstm-based transport mode detection. In *Ieee international conference on pervasive computing and communications workshops (percom workshops)* (pp. 107–112).

Ashqar, H. I., Elhenawy, M., Rakha, H. A., Almannaa, M., & House, L. (2022). Network and station-level bike-sharing system prediction: A san francisco bay area case study. *Journal of Intelligent Transportation Systems*, *26*(5), 602–612.

Ballı, S., & Sağbaş, E. A. (2018). Diagnosis of transportation modes on mobile phone using logistic regression classification. *IET Software*, *12*(2), 142–151.

Biancat, J., Brighenti, C., & Brighenti, A. (2014). Review of transportation mode detection techniques. *EAI Endorsed Transactions on Ambient Systems*, *1*(4).

Byon, Y.-J., Ha, J. S., Cho, C.-S., Kim, T.-Y., & Yeun, C. Y. (2017). Real-time transportation mode identification using artificial neural networks enhanced with mode availability layers: A case study in dubai. *Applied Sciences*, *7*(9), 1–17.

Carpineti, C., Lomonaco, V., Bedogni, L., Di Felice, M., & Bononi, L. (2018). Custom dual transportation mode detection by smartphone devices exploiting sensor diversity. In *International conference on pervasive computing and communications workshops (percom workshops)* (pp. 367–372).

Catalán, H., Lobel, H., & Herrera, J. C. (2022). An automatic methodology to measure drivers' behavior in public transport. *Journal of Intelligent Transportation Systems*, 1–15.

Coroamă, V. C., Türk, C., & Mattern, F. (2019). Exploring the usefulness of bluetooth and wifi proximity for transportation mode recognition. In *Acm international joint conference on pervasive and ubiquitous computing and international symposium on wearable computers* (pp. 37–40).

Dawson, R. (2011). How significant is a boxplot outlier? *Journal of Statistics Education*, *19*(2), 143–162.

Delli Priscoli, F., Giuseppi, A., & Lisi, F. (2020). Automatic transportation mode recognition on smartphone data based on deep neural networks. *Sensors*, *20*(24), 1–16.

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.

Fang, S.-H., Liao, H.-H., Fei, Y.-X., Chen, K.-H., Huang, J.-W., Lu, Y.-D., & Tsao, Y. (2016). Transportation modes classification using sensors on smartphones. *Sensors*, *16*(8), 1–15.

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*(2), 131–139.

Hadjidimitriou, N. S., Cantelmo, G., & Antoniou, C. (2022). Machine learning for activity pattern detection. *Journal of Intelligent Transportation Systems*, 1–15.

Hasan, R. A., Irshaid, H., Alhomaidat, F., Lee, S., & Oh, J.-S. (2022). Transportation mode detection by using smartphones and smartwatches with machine learning. *KSCE Journal of Civil Engineering*, *26*(8), 3578–3589.

He, X., Zhao, K., & Chu, X. (2021). Automl: A survey of the state-of-the-art. *Knowledge-Based Systems*, *212*, 1–27.

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.

Kamalian, M., Ferreira, P., & Jul, E. (2022). A survey on local transport mode detection on the edge of the network. *Applied Intelligence*, 1–30.

Kostrzewska, M., & Macikowski, B. (2017). Towards hybrid urban mobility: Kick scooter as a means of individual transport in the city. *IOP Conference Series: Materials Science and Engineering*, *245*, 1–9.

Liang, X., & Wang, G. (2017). A convolutional neural network for transportation mode detection based on smartphone platform. In *International conference on mobile ad hoc and sensor systems (mass)* (pp. 338–342).

Liang, X., Zhang, Y., Wang, G., & Xu, S. (2020). A deep learning model for transportation mode detection based on smartphone sensing data. *IEEE Transactions on Intelligent Transportation Systems*, *21*(12), 5223-5235.

Liu, X. (2022). Glmlp-trans: A transportation mode detection model using lightweight sensors integrated in smartphones. *Computer Communications*, *194*, 156-166.

Lorintiu, O., & Vassilev, A. (2016). Transportation mode recognition based on smartphone embedded sensors for carbon footprint estimation. In *International conference on intelligent transportation systems (itsc)* (pp. 1976–1981).

McKenzie, G. (2020). Urban mobility in the sharing economy: A spatiotemporal comparison of shared mobility services. *Computers, Environment and Urban Systems*, *79*, 1–10.

Nham, B., Siangliulue, K., & Yeung, S. (2008). Predicting mode of transport from iphone accelerometer data. *Machine Learning Final Projects, Stanford University*.

Oeschger, G., Carroll, P., & Caulfield, B. (2020). Micromobility and public transport integration: The current state of knowledge. *Transportation Research Part D: Transport and Environment*, *89*, 1-21.

Osoba, M. Y., Rao, A. K., Agrawal, S. K., & Lalwani, A. K. (2019). Balance and gait in the elderly: A contemporary review. *Laryngoscope investigative otolaryngology*, *4*(1), 143–153.

Prelipcean, A. C., Gidófalvi, G., & Susilo, Y. O. (2017). Transportation mode detection–an in-depth review of applicability and reliability. *Transport reviews*, *37*(4), 442–464.

Prentow, T. S., Blunck, H., Kjærgaard, M. B., & Stisen, A. (2015). Towards indoor transportation mode detection using mobile sensing. In *International conference on mobile computing, applications, and services* (pp. 259–279).

Roy, A., Fuller, D., Nelson, T., & Kedron, P. (2022). Assessing the role of geographic context in transportation mode detection from gps data. *Journal of Transport Geography*, *100*, 103330.

Sadeghian, P., Håkansson, J., & Zhao, X. (2021). Review and evaluation of methods in transport mode detection based on gps tracking data. *Journal of Traffic and Transportation Engineering (English Edition)*, *8*(4), 467–482.

Sankaran, K., Zhu, M., Guo, X. F., Ananda, A. L., Chan, M. C., & Peh, L.-S. (2014). Using mobile phone barometer for low-power transportation context detection. In *Acm conference on embedded network sensor systems* (pp. 191–205).

Shah, R. C., Wan, C.-y., Lu, H., & Nachman, L. (2014). Classifying the mode of transportation on mobile phones using gis information. In *Acm international joint conference on pervasive and ubiquitous computing* (pp. 225–229).

Sharma, A., Singh, S. K., Udmale, S. S., Singh, A. K., & Singh, R. (2021). Early transportation mode detection using smartphone sensing data. *IEEE Sensors Journal*, *21*(14), 15651-15659.

Shiv, A. (2018). *Analysis of last mile transport pilot: Implementation of the model and its adaptation among local citizens* (Unpublished master's thesis). Aalto University, Finland.

Su, X., Yao, Y., He, Q., Lu, J., & Tong, H. (2017). Personalized travel mode detection with smartphone sensors. In *2017 ieee international conference on big data (big data)* (pp. 1341–1348).

Taia Alaoui, F., Fourati, H., Kibangou, A., Robu, B., & Vuillerme, N. (2022). Urban transportation mode detection from inertial and barometric data in pedestrian mobility. *IEEE Sensors Journal*, *22*(6), 4772–4780.

Taia Alaoui, F., Fourati, H., Vuillerme, N., Kibangou, A., Robu, B., & Villemazet, C. (2020). *Tmd-captimove.* 2020 via Perscido-Grenoble-Alpes.

Wang, Luo, H., Zhao, F., & Qin, Y. (2021). Combining residual and lstm recurrent networks for transportation mode detection using multimodal sensors integrated in smartphones. *IEEE Transactions on Intelligent Transportation Systems*, *22*(9), 5473-5485.

Wang, B., Vu, H. L., Kim, I., & Cai, C. (2022). Short-term traffic flow prediction in bike-sharing networks. *Journal of Intelligent Transportation Systems*, *26*(4), 461–475.

Wang, L., Gjoreski, H., Ciliberto, M., Mekki, S., Valentin, S., & Roggen, D. (2019). Enabling reproducible research in sensor-based transportation mode recognition with the sussex-huawei dataset. *IEEE Access*, *7*, 10870-10891.

Wang, L., Gjoreskia, H., Murao, K., Okita, T., & Roggen, D. (2018). Summary of the sussex-huawei locomotion-transportation recognition challenge. In *Acm international joint conference and international symposium on pervasive and ubiquitous computing and wearable computers* (pp. 1521–1530).

Wang, L., & Roggen, D. (2019). Sound-based transportation mode recognition with smartphones. In *Ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 930–934).

Xiao, G., Juan, Z., & Zhang, C. (2015). Travel mode detection based on gps track data and bayesian networks. *Computers, Environment and Urban Systems*, *54*, 14–22.

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. *Proceedings of the VLDB Endowment*, *7*(13), 1429–1440.

Zarei Yazd, M., Taheri Sarteshnizi, I., Samimi, A., & Sarvi, M. (2022). A robust machine learning structure for driving events recognition using smartphone motion sensors. *Journal of Intelligent Transportation Systems*, 1–15.

Zhao, H., Hou, C., Alrobassy, H., & Zeng, X. (2019). Recognition of transportation state by smartphone sensors using deep bi-lstm neural network. *Journal of Computer Networks and Communications*, *2019*.

Zheng, Y., Liu, L., Wang, L., & Xie, X. (2008). Learning transportation mode from raw gps data for geographic applications on the web. In *Proceedings of the 17th international conference on world wide web* (pp. 247–256).

Zhou, C., Jia, H., Gao, J., Yang, L., Feng, Y., & Tian, G. (2017). Travel mode detection method based on big smartphone global positioning system tracking data. *Advances in Mechanical Engineering*, *9*(6).

Zimmer, L., Lindauer, M., & Hutter, F. (2021). Auto-pytorch: multi-fidelity metalearning for efficient and robust autodl. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(9), 3079–3090.

Zuo, T., Wei, H., Chen, N., & Zhang, C. (2020). First-and-last mile solution via bicycling to improving transit accessibility and advancing transportation equity. *Cities*, *99*(102614), 1–14.

Appendix A. Model design: Random Forest (RF) and Feed-forward Neural Network (ANN)

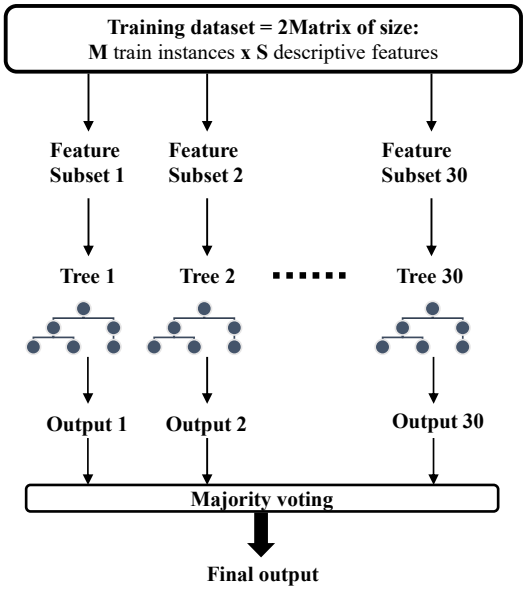


Figure 1.: Random Forest

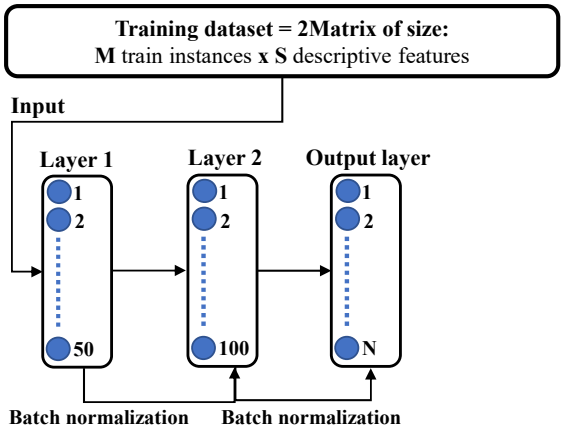


Figure 2.: ANN architecture

Appendix B. Model design: Long-Short Term Memory (LSTM) network and Convolutional Neural Network (CNN)

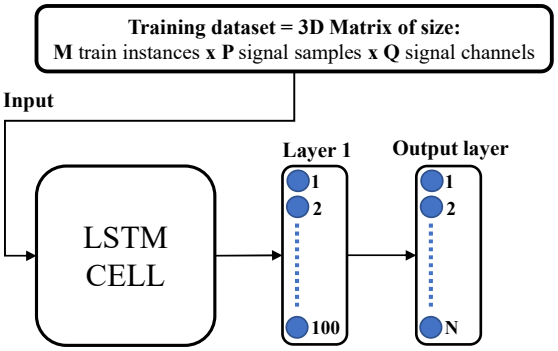


Figure 1.: LSTM architecture

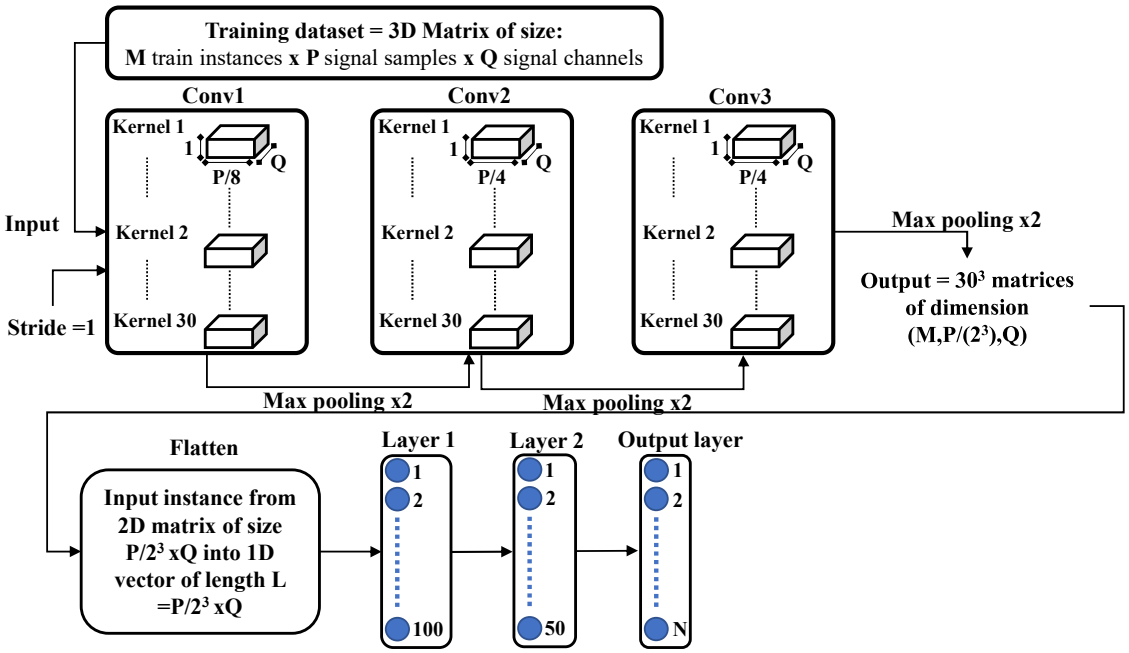


Figure 2.: CNN architecture

Appendix C. Confusion matrices for the first and second scenario

Table 1.: Confusion matrices for the first scenario

Foot		W	B	KS
U1	W	1934	35	121
	B	0	1012	147
	KS	121	250	1147
Overall Accuracy				**85.86%**
F1-Score				**84.41%**

Waist		W	B	KS
	W	1811	89	66
	B	28	266	0
	KS	186	1020	53
				60.53%
				28.13%

Pocket		W	B	KS
	W	1603	0	26
	B	35	0	464
	KS	342	0	1989
				80.56%
				85.71%

Foot		W	B	KS
U2	W	1542	0	25
	B	0	1286	956
	KS	14	402	2480
Overall Accuracy				**79.16%**
F1-Score				**79.59%**

Waist		W	B	KS
	W	3361	75	22
	B	2	450	392
	KS	33	1144	3112
				80.58%
				65.40%

Pocket		W	B	KS
	W	3489	32	2
	B	0	1099	285
	KS	71	2231	3103
				74.58%
				68.54%

Foot		W	B	KS
U3	W	0	0	9
	B	0	0	1324
	KS	0	0	2410
Overall Accuracy				**64.39%**
F1-Score				**78.34%**

Waist		W	B	KS
	W	0	0	243
	B	0	0	701
	KS	0	0	2326
				71.13%
				83.13%

Pocket		W	B	KS
	W	0	0	42
	B	0	0	98
	KS	0	0	3464
				96.12%
				98.02%

Foot		W	B	KS
U4	W	1934	35	121
	B	0	1012	147
	KS	121	250	1147
Overall Accuracy				**85.86%**
F1-Score				**84.41%**

Waist		W	B	KS
U5	W	2085	16	341
	B	38	1426	1828
	KS	28	191	1877
				68.81%
				68.49%

Pocket		W	B	KS
	W	2186	0	58
	B	7	1424	514
	KS	111	262	1883
				85.23%
				84.47%

Foot		W	B	KS
U5	W	482	0	0
	B	0	1219	535
	KS	22	173	1920
Overall Accuracy				**83.22%**
F1-Score				**86.03%**

Waist		W	B	KS
	W	6215	0	4
	B	765	169	787
	KS	8290	1252	1408
				41.25%
				23.71%

Pocket		W	B	KS
	W	231	40	67
	B	0	1099	1378
	KS	2	1728	3623
				60.64%
				61.39%

Table 2.: Confusion matrices of the second scenario

Foot		W	B	PT	KS
U1	W	1339	0	0	12
	B	0	837	1	205
	PT	397	326	4311	147
	KS	319	134	31	1024
Overall Accuracy					**82.45%**
F1-Score					**77.63%**

Waist	W	B	PT	KS
W	626	0	0	6
B	192	279	12	18
PT	13	424	4242	8
KS	1189	672	71	87
				66.77%
				44.81%

Pocket	W	B	PT	KS
W	930	0	0	40
B	344	0	158	821
PT	23	0	4093	290
KS	683	0	81	1328
				72.24%
				53.71%

Foot		W	B	PT	KS
U2	W	1481	0	6	1
	B	1	1506	16	1370
	PT	27	100	4406	602
	KS	42	81	29	1488
Overall Accuracy					**79.61%**
F1-Score					**78.36%**

Waist	W	B	PT	KS
W	3095	10	34	0
B	0	453	5	113
PT	7	307	4718	2714
KS	289	900	173	699
				66.32%
				58.67%

Pocket	W	B	PT	KS
W	3021	5	16	0
B	0	1375	16	121
PT	167	234	47502214	
KS	367	1748	63	1155
				67.54%
				64.92%

Foot		W	B	PT	KS
U3	W	0	0	0	4
	B	0	0	0	2837
	PT	0	0	0	39
	KS	0	0	0	863
Overall Accuracy					**23.06%**
F1-Score					**37.47%**

Waist	W	B	PT	KS
W	0	0	0	5
B	0	0	0	1412
PT	0	0	0	807
KS	0	0	0	1046
				31.99%
				48.47%

Pocket	W	B	PT	KS
W	0	0	0	0
B	0	0	0	696
PT	0	0	0	176
KS	0	0	0	2732
				75.80%
				86.24%

Foot		W	B	PT	KS
U4	W	1027	0	0	23
	B	1	1725	16	1566
	PT	44	285	3801	259
	KS	89	43	0	2174
Overall Accuracy					**78.96%**
F1-Score					**79.65%**

Waist	W	B	PT	KS
W	1946	0	0	12
B	26	1228	14	603
PT	54	314	4503	1757
KS	124	90	47	1674
				75.46%
				75.32%

Pocket	W	B	PT	KS
W	1886	0	0	20
B	41	1273	40	481
PT	40	336	3369	735
KS	338	77	31	1220
				78.36%
				76.54%

Foot		W	B	PT	KS
U5	**W**	501	1	30510	0
	B	0	559	391654	2
	PT	0	0	628957	0
	KS	2	832	102790	2453
Overall Accuracy					**54.61%**
F1-Score					**19.63%**

Waist	W	B	PT	KS
W	12139	14	4098	176
B	825	1206	157093	1248
PT	188	20	716323	3
KS	2111	180	154584	772
				69.50%
				40.25%

Pocket	W	B	PT	KS
W	228	150	17777	611
B	0	778	187677	700
PT	0	0	519417	2
KS	5	1940	329722	3756
Overall Accuracy				**49.32%**
F1-Score				**17.85%**