



HAL
open science

Urban Transportation Mode Detection From Inertial and Barometric Data in Pedestrian Mobility

F. Taia Alaoui, Hassen Fourati, Alain Kibangou, Bogdan Robu, N. Vuillerme

► **To cite this version:**

F. Taia Alaoui, Hassen Fourati, Alain Kibangou, Bogdan Robu, N. Vuillerme. Urban Transportation Mode Detection From Inertial and Barometric Data in Pedestrian Mobility. *IEEE Sensors Journal*, 2022, 22 (6), pp.4772-4780. 10.1109/JSEN.2021.3065848 . hal-04055885

HAL Id: hal-04055885

<https://hal.science/hal-04055885>

Submitted on 4 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Urban Transportation Mode Detection From Inertial and Barometric Data in Pedestrian Mobility

F. Taia Alaoui¹, H. Fourati¹, A. Kibangou¹, B. Robu¹, and N. Vuillerme

Abstract—This paper presents a novel urban transportation mode detection (TMD) method based on inertial and barometric data. To the best of the authors knowledge, this is the first time that a 3-axis accelerometer, a 3-axis gyroscope, and a barometer are combined for TMD purposes. One contribution was to build and share an optimized TMD dataset in terms of duration (44 hours), number of participants (34), sensor placements (on-foot, waist-attached and in the trouser's pocket) and considered transportation modes (7). To infer transportation mode information, we tackle two classification scenarios: still, walk, bike, tramway, and bus, on the one hand, and still, walk, bike, elevator, stairs and public transport on the other hand. One major finding was that TMD classification accuracy has been widely overestimated by adopting the familiar randomized cross validation approach. According to the latter, the accuracy of our non-optimized classifiers ranged from 90% to 99%. However, few of them were accurate when the test samples were drawn from data of five unknown subjects. By selecting features in a personalized manner and by studying the importance of the placement of sensors, we realized that the latter significantly impact the performance of TMD models. Thus, we present here two robust TMD classifiers that show respectively an average accuracy of 75.63% for the first scenario and 79.41% for the second scenario using a foot-mounted sensor

Index Terms—Transportation mode detection, inertial sensors, barometer, indoor-outdoor scenarios, activities recognition and classification, data collection.



I. INTRODUCTION

TRANSPORTATION mode information is valuable to several sectors such as transport and traffic planning, location-based services, and healthcare applications. In the last decade, many studies have considered solutions that identify transportation modes using Global Positioning System (GPS) [1] or mobile networks [2]. These networks have been sometimes fused with Geographical Information Systems (GIS) [3] or inertial sensors [4] to achieve higher accuracies.

Manuscript received October 24, 2020; revised February 26, 2021; accepted March 8, 2021. Date of publication March 12, 2021; date of current version March 14, 2022. This work was supported in part by the Project CAPTIMOVE by the IDEX of University Grenoble Alpes [Initiatives de Recherche Stratégiques (IRS)] under Grant ANR-15-IDEX-02 and in part by the French National Research Agency in the Framework of the Investissements d'avenir Program under Grant ANR-10-AIRT-05. The associate editor coordinating the review of this article and approving it for publication was Dr. Kyle O'keefe. (Corresponding author: H. Fourati.)

The authors are with GIPSA-Lab, Grenoble INP, Inria, CNRS, University Grenoble Alpes, 38000 Grenoble, France (e-mail: fadoua.taia-alaoui@gipsa-lab.fr; hassan.fourati@gipsa-lab.fr; alain.kibangou@gipsa-lab.fr; bogdan.robustu@gipsa-lab.fr; nicolas.vuillerme@univ-grenoble-alpes.fr).

For example, GPS and GIS were fused in [5] to recognize user's travel mode. Others have used Bluetooth or Wireless Fidelity (WIFI) [6]. Yet, these solutions require GPS availability, or infrastructure deployment and maintenance in the case of other terrestrial wireless networks. On the other hand, it is established that the most energy consuming technology is GPS, followed by Global System for Mobile communications (GSM) which in turn is followed by WIFI [7]. Therefore, energy efficient sensors have been suggested instead. They include accelerometers [8], gyroscopes [9], barometers [10], magnetometers [11] and microphones [12], [13]. To enhance TMD systems, different combination sets of these sensors were proposed. For example, accelerometer and GPS [14] and accelerometer and mobile network data [15] were fused to infer transportation mode information. More exhaustive combination sets were suggested such as the combination of an accelerometer, a gyroscope and a microphone in [12], GPS, an accelerometer, a gyroscope, a magnetometer, and GIS in [16], or a light sensor, a barometer, an accelerometer, a gyroscope and a magnetometer in [10].

Beside the huge amount of the technologies used to infer transportation modes, there is an equivalent diversity

in the state-of-the-art regarding considered classifications. In the beginnings of TMD research, studies distinguished only between vehicle (i.e. motorized) and on-foot modes [17]. Today, almost all motorized modes have been at least once considered independently. Indeed, car [18], bus [14], motorcycle [19], truck [16], plane [20] and boat [16] have all been individually classified. Some studies have even fine-tuned some of these modes, like [21] who separated national buses from urban buses, and others [7], [22], [23] who classified separately High Speed Rail (HSR), train and metro (also called subway in other studies [24]). Similarly, tramway, train and subway/metro were separated in [19], while human-powered and electric bikes (E-bike) were distinguished in [1]. In [20], a classification from the perspective of infrastructure was adopted by introducing a class called ‘Rail’ and a class called ‘Road’ in addition to plane and other on-foot motion modes. As a result, a cross-study comparison seems inadequate due to the large discrepancies between these scenarios. In order to alleviate this problem, the TMD research community could share their datasets to enable benchmarking and reproducibility [25]. Besides, the evaluation criteria of TMD models should be standardized as important biases might be introduced whether new separate subjects are involved in model testing or not.

In this work, the main contributions can be summarized as follows: First, we provide a new TMD dataset made public and shared online¹ for benchmarking and reproducibility. We limit the spatial scale of our experiments to urban spaces and public transport (tramway and bus). Thus, car, train, national bus, and HSR are excluded. Second, we consider 3 sensor placements, on-foot, in the trouser’s pocket, and waist-worn, to study the impact of each sensor placement on the classification results. Third, we demonstrate that a major part of previous TMD works overestimated the classification accuracy by not considering cross-subject variability [26], [27], whereas few others used data of new subjects to test the TMD models and obtained lower accuracy rates [24], or had to constrain the use of the sensors to some specific position and orientation [9]. We follow in this paper both methods and demonstrate that overestimated accuracy rates result from the first evaluation method, which we call randomized cross validation. Finally, we provide two optimal classification models for two classification scenarios. These models are validated on 5 test subjects that are isolated, one by one, from the train subjects. The classification performance is given as a 2D vector comprising an average accuracy and a standard deviation, which represents the ability of the model to deal with cross-subject variability, and could be seen as a robustness index.

The paper is organized as follows: Section II presents the main past TMD studies, since 2017. Section III describes the data collection. Section IV details the followed methodology to construct the training database. Section V describes the classification approach. Section VI presents the results. Section VII provides a discussion of these results, while section VIII is a brief conclusion of this work.

II. LITERATURE REVIEW

A. Study Selection

Table I provides the main recent (≥ 2017) TMD studies and datasets. Indeed, studies earlier than 2017 have produced many of the datasets used after 2017 with more sophisticated models. Besides, an important part of past studies were reliant either on GPS or WIFI. In this paper, the attention is directed towards solutions that fit into the range of GPS and infrastructure free systems. Therefore, considered studies in this literature review utilize mainly inertial sensors, which are eventually fused with non-inertial sensors such as microphones and magnetometers in order to improve the classification accuracy. The datasets of interest here are either public or private. Some datasets may also include GPS data [24], but the models developed in the chosen references can also be functional without GPS. Otherwise, a study is not cited because the focus here is mainly directed towards inertial sensors-based solutions for more energy efficiency. In the same way, studies in which GIS layers are necessary such as [16], or WIFI, are discarded.

B. Literature Table Construction

In Table I, the abbreviation NM corresponds to ‘Not Mentioned’, meaning that the information is lacking in the study. The table is organized in eight columns. The seventh column indicates the evaluation method used to test the classifier. The last column gives the classification overall accuracy. This couple of data, in addition to the used sensors, should be considered together while comparing performances of different solutions. In fact, in most cases, all data from all participants are merged and a K-fold cross-validation process is followed to estimate the overall accuracy. According to this method, the model ability to generalize to new subjects is not evaluated and the classification accuracy is overestimated. In few studies, one or more test subjects were left for evaluation (column 7 = Yes). In this case, performance rates are lower, which is more realistic as the trained models cannot handle the specific patterns relative to the new test subjects. Hence, the classification accuracy is higher or lower depending on the chosen test subject, which in average results in lower accuracy.

III. DATA COLLECTION

A. Experimental Protocol

Prior to any experiment, a data use and participation agreement was provided by every participant. In total, 34 healthy adults voluntarily participated to the study, 14 women and 20 men aged from 18 to 50 years. The experiments were conducted during workdays at different hours of the day (8AM-6PM), including rush hours (8-9AM, 12-2AM, 4-6PM), under rainy, sunny, and cloudy weather conditions. In total, the data collection lasted around 3 months. Nearly 44 hours of data were recorded. For logistical reasons, routes taken during experiments were fixed for the following 4 transportation/locomotion modes: tramway, bus, elevators, and stairs. Each participant was accompanied during the experiment by the person in charge of labeling the data. Then, for the remaining modes: still, walking, and biking, participants had

¹<https://perscido.univ-grenoble-alpes.fr/datasets/DS310>

TABLE I
MAIN GPS AND GIS-FREE TMD WORKS SINCE 2017

Reference & year	Sample size	Duration (hour)	Classified modes	Data	Sensor placements	Separate test/train subjects	Accuracy
[26], 2019 using the dataset of [24], 2019	1	NM	Still, Walk, Bike, Run, Bus, Car, Train, Subway	Acc, Gyr, Mag, Bar	Hip, torso, and bag for training. Hand for testing	No	67.5%
[27], 2019 using the dataset of [7], 2014	NM (\neq [7])	211.82	Still, Walk, Bike, Run, Bus, Car, Train, Metro, High Speed Rail, Motorcycle	Acc, Gyr, Mag	NM	No	96.82%
[28], 2019	NM	NM	Still, Walk, Bike, Bus, Car, Train, Subway	Acc	No requirement	No	95%
[29], 2019 using the dataset of [30], 2015	10	25	Walk, Bike, Run, Bus, Car, Bike	Acc, Gyr	No requirement	No	97%
[9], 2019	11	NM	Still, Walk, Bike, Run, Bus, Subway	Acc, Gyr	Waist+fixed orientation	Yes	93%
[24], 2019 same dataset as [31], 2018	3	703	Still, Walk, Bike, Run, Bus, Car, Train, Subway	Acc, Gyr	4 placements	Yes	57.5%
[12], 2018	13	32	Still, Walk, Bus, Car, Train	Acc, Gyr, Microphone	No requirement	Yes	81%
[32], 2018	58	~500	Bus, Car, Train, Subway	Acc, Gyr, Mag, Bar, Base station signal strength,	No requirement	No	97%
[33], 2018	NM	8	Walk, Bike, Run, Bus, Car	Acc, Gyr	NM	No	99.5%
[11], 2018 using the dataset of [20], 2016	22	225	Still, Walk, Bike, Run, Road, Rail, Plane, Other	Acc, Mag	No requirement	No	94%
[10], 2017	12	NM	Walk, Bike, Jogging, Bus, Car, Subway	Acc, Gyr, Mag, Bar, Gravity, Light, Gender, Season, City, Phone model	No requirement	No	70-95%
[22], 2017 using part of the data of [7], 2014	NM	>1000	Still, Walk, Bike, Run, Vehicle	Acc, Gyr, Mag	NM	No	95%
[8], 2017	NM	2/mode	Still, Walk, Bike, Bus, Car, Train, Subway	Acc	No requirement	No	95%

no special requirements. In fact, some of them used their own bike and felt comfortable enough to make longer trips, while others who seldom used bikes made shorter trips. However, all participants were asked to make a ride lasting more than a minimum duration that was fixed for each mode. For example, each participant had to ride a bike for at least

5min. Overall, at least 20 participants have collected data for each mode. More details about the number of participants and the accumulated time-duration for each transportation and locomotion mode are described in Fig.1 and 2.

As compared with other TMD databases presented in Table I, the current one is the second most important with

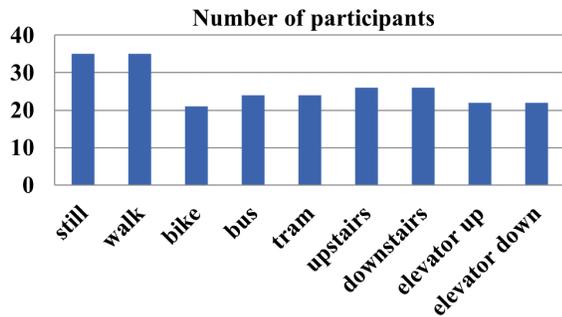


Fig. 1. Number of participants per transportation mode.

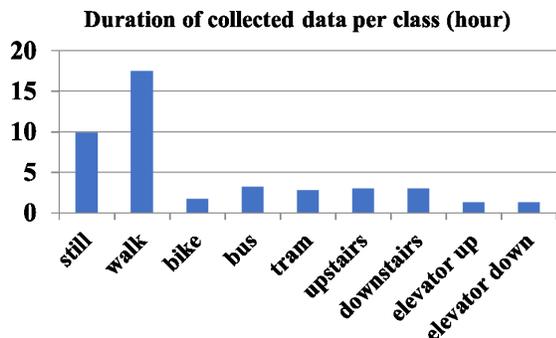


Fig. 2. Total time duration corresponding to each class.

regard to the number of participants (34). It also has more balance between the amount of data in terms of duration and the number of considered classes (except for static and walking, which is representative of the actual time distribution of daily activities). It provides data of 3 sensor placements, which allows to study the sensors placements impact on the classification. As compared with the studies of Table I, there are only 3 studies that specify the placement of the sensors (row 1, 5 and 6 of Table I). Additionally, it covers indoor activities, which are totally absent from the studies listed in Table I and are more generally rare in TMD studies that depend on additional information and sensors.

B. Hardware

Data were collected using 3 inertial measurement units (IMUs) that are designed for human motion analysis purposes. The IMUs belong to the Gaitup Physilog5 series.² They embed a 3-axis accelerometer, a 3-axis gyroscope and a barometer each. The sensors manufacturer is STMicroelectronics. The accelerometer sensitivity is up to 16g where g is the gravity constant, while the gyroscope's sensitivity is up to 2000°/s. Finally, the barometer has a sensitivity range between 260 and 1260 hPa. The raw sampling frequency of accelerometers and gyroscopes signals is 128 Hz, while barometers deliver measurements at a frequency of 64 Hz. The inertial sensors were synchronized inside each IMU, meaning that acceleration and angular rate are natively synchronized, while pressure had to be interpolated at the accelerometer/gyroscope frequency. Data used in this study were sub-sampled at 32 Hz to lower the computational cost.

²<https://research.gaitup.com/physilog/>

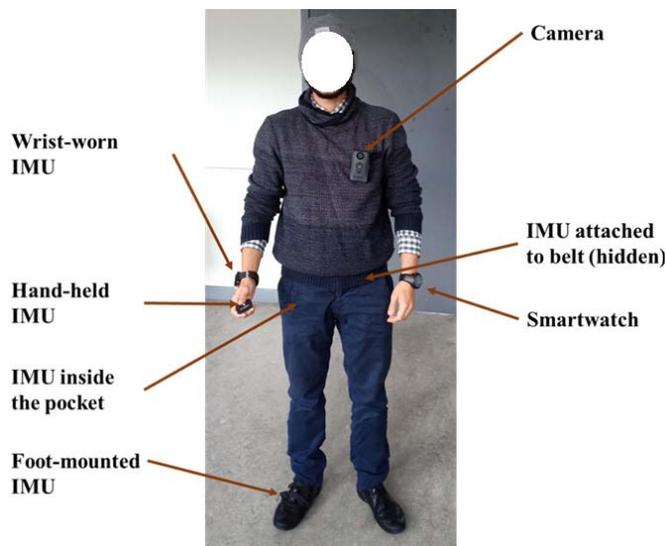


Fig. 3. Placement of the sensors.

The 3 used IMUs were respectively attached to the foot, to the waist, and placed in the trousers pocket. A fourth IMU was held in hand by the participants and a fifth IMU was worn in the wrist. The data from these two last IMUs are not exploited in this work but are available in the released dataset. The placement of the IMUs is illustrated in Fig. 3. A camera was used for visual checking and a smartwatch was used to measure heart rate. In this paper, heart rate is left for future work, but the collected measurements are provided in the released dataset.

IV. PREPROCESSING AND TRAINING DATASET PREPARATION

A. Preprocessing of the Collected Signals

Signals have been preprocessed before feature computation as follows:

- Acceleration: Boxplots were computed for signal data frames that are 1 s long without overlapping. Inside each data frame, the upper and lower outliers were respectively replaced by the maximum and minimum value of the boxplot. This preprocessing strategy aims at deleting isolated peaks that occur due to undesirable motion that is independent from the transportation mode. The peak detection has been first applied to each axis, and then the norm of the signal was computed.
- Angular rate: The same process such as for acceleration has been applied to angular rate.
- Atmospheric pressure: Due to high frequency noise in measured pressure, we have smoothed the signal by computing its average envelope.

B. Signal Segmentation

After we analyzed collected signals, the periods corresponding to each transportation mode/class have been segmented according to a sliding window of 3 s. The average length of the segmenting window in the state-of-the-art is around 5 s [24]. However, because we introduced some indoor activities such as elevators and stairs, a 5 s window seemed to be too long as we observed that moving from one floor to another took

TABLE II
SET OF INPUT FEATURES BEFORE SELECTION/EXTRACTION

Signal channel	Features
Acceleration norm (norm acc), Gradient of the norm of acceleration (grad acc), Integral of the norm of acceleration (int acc), Angular rate norm (norm gyr), Gradient of the norm of angular rate (grad gyr), Integral of the norm of angular rate (int gyr)	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
Smoothed pressure (pres)	Spectral centroid, Spectral Spread, Number of zero crossings after scaling, 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
Gradient of smoothed pressure (grad pres) and integral of smoothed pressure (int pres)	Mean, Min, Max, Standard deviation, Main frequency component, Power of the main frequency component, Spectral centroid, Spectral spread, Number of zero crossings

less than 3 s using the stairs or the elevators. On the other hand, choosing 1 s segmenting window, as suggested in some studies [29], [10], did not lead to the best performance.

C. Feature Computation

Table II provides the details of computed features in both frequency and time-domains for each signal channel. They have been chosen based on a physical interpretation of the signals as well as from state-of-the-art studies based on the spectral energy in some specific sub-bands [24] or the variance/standard deviation [12]. In summary, the same features were computed for the channels based on acceleration and angular rate as they behave quite similarly. However, maximum and minimum were not computed for these signals in order to dismiss isolated peaks that are not related to the transportation modes. For pressure-based signals, the mean, maximum and minimum of smoothed pressure were not computed because they are location-specific, however, this limitation does not exist for the integral and gradient channels. Therefore, more time-domain features were computed for these two last channels.

V. CLASSIFICATION APPROACH

From the review of the literature, the comparison of different machine learning models, such as Naive Bayes, Support Vector Machines (SVM), or Decision trees, has shown that Random

Forest (RF) outperforms the other models [20], [12], [25]. Besides, deep neural networks have also been used in TMD studies, including feed-forward neural networks (ANN) [22], [12], convolutional neural networks (CNN) [8], and long-short term memory (LSTM) neural networks [27], [32]. However, unlike RF where the hyper-parameters (the number of trees, subset of features in each decision tree and the split criterion) do not require a fine tuning, neural networks are more sensitive to the tuning of both the hyper-parameters related to the network structure and those related to the training process. Therefore, up to the present, neural networks have been more subject to overfitting and have proven less efficient than machine learning algorithms when tested on unknown TMD datasets [25]. Given these considerations, multiple models were tested on test subjects who were not involved in training. These models include RF, ANN, CNN, and LSTM. By testing these models on new data, we were able to verify whether the results of this study are compliant with the results expected from the state-of-the-art regarding the supremacy of RF.

In order to optimize the feature space, different feature selection and extraction methods were realized. We used Principal Component Analysis (PCA) feature extraction, Kendall's Tau (KT)-based feature selection, and mutual information (MI)-based feature selection. With these 3 strategies, 3 different datasets were constructed and added to the global dataset that includes all computed features (section IV.C). In total, this produced 4 different datasets. The 4 were used for training with RF and ANN. As for CNN and LSTM, the training was made with the pre-processed signals so that the data time-structure was preserved inside training samples. Finally, we developed a customized method to select the most relevant signal channels through a process of several classifications with each signal channel taken alone, and then validated on 5 different test subjects. This allowed us selecting the most relevant channels, which are obviously the channels that produce the most accurate classification results in average. The selected channels were then combined in one new subset and multiple classification tests were realized by varying the number of used channels. By removing a channel, we could see if the results were improved. In that case, the channel was removed, otherwise it was kept.

VI. RESULTS

In the following sections, accuracy rates are expressed either in terms of overall accuracies (OA) or Macro-averaged F1-Scores, which is noted F1-Score for simplicity. In fact, the OA is a popular accuracy metric that is biased towards the most represented class in case of highly imbalanced datasets. As this is the case in this study, the macro-averaged F1-Score has been adopted together with the OA in order to provide a more complete and unbiased evaluation of the classification results. In fact, with the F1-Score, each class has an equal weight and the global macro-averaged F1-Score is not impacted by the class distribution. These evaluation metrics can be found in detail in [34]. The classification results are presented for the two following scenarios:

- Still, walk, bike, tramway, bus
- Still, walk, bike, elevators, stairs, public transport (PT)

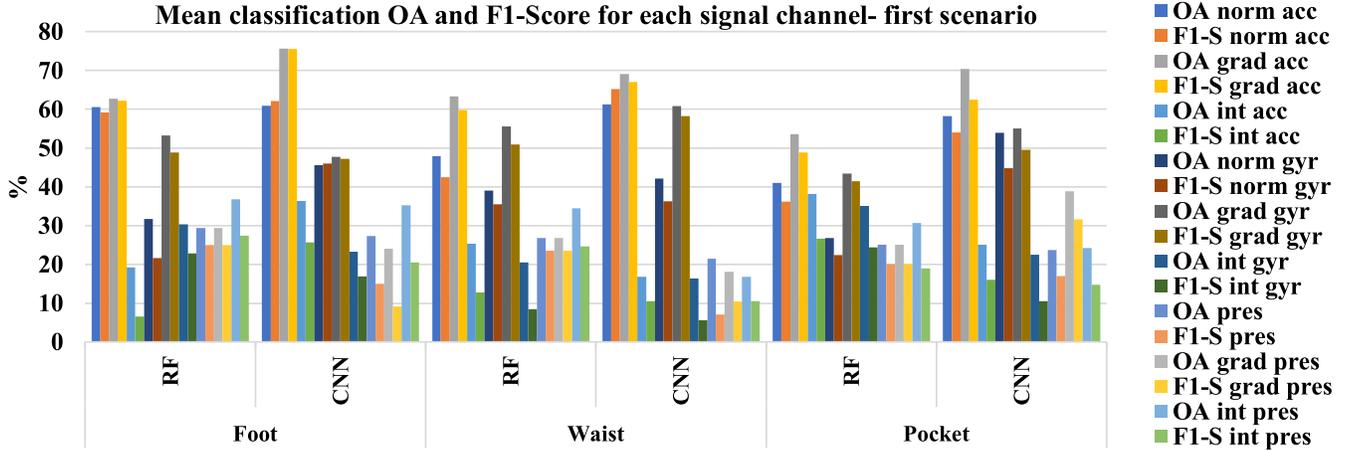


Fig. 4. Channel by channel classification for the first scenario.

A. Summary of the Model Selection Results

According to the sub-mentioned classification approach, low performances were achieved using the first generic feature selection and extraction methods (PCA, KT, and MI) used with ANN and RF. Besides, using all computed features, even with RF, did not lead to satisfying results. Because the tests are numerous and time-consuming, the results of the previously mentioned methods were not performed for all test subjects and their results are not provided in this paper. Yet, note that testing all of them provided us with an important first selection result. In fact, after discarding the least efficient algorithms relative to this classification problem, we could focus only on the relevant models which are presented in details in this paper. Unlike the generic feature optimization methods, the customized channel selection strategy (see section V) was more fruitful. Indeed, it has shown a higher classification performance together with less signal channels needed to perform the classification, which improves the computational efficiency. Yet, the supremacy of RF was not confirmed in this work because the highest classification accuracy was obtained using a CNN. As the model selection is also dependent on the chosen training features and signal channels, we found that the best results were obtained with one signal-channel used for the first scenario, which is the gradient of acceleration, and two channels for the second scenario, which are the gradients of pressure and of acceleration. More importantly, these results show that with an accelerometer and a barometer, transportation modes can be recognized without using a gyroscope, which is known to be more energy consuming than an accelerometer or a barometer [7]. In order to visualize the importance of each signal, next sections will provide the detailed results of the channel by channel classification results using RF and CNN.

B. First Scenario Classification Results

1) *Channel by Channel Classification Results With RF and CNN:* Fig. 4 shows the pairs of OA and F1-Score relative to each signal channel for RF and CNN and for the 3 sensor placements. The variables in these figures are explained in the first column of Table II. From these results, it is observed that

the most important channels, sorted decreasingly based on the F1-Score value, are first the gradient of acceleration, second, the norm of acceleration, third, the gradient of angular rate, and fourth, the angular rate norm. It is interesting to notice that this order is the same for all sensor placements, and for all algorithms. Additionally, the classification accuracy obtained with the two first channels (acceleration norm and gradient of acceleration) is systematically higher for the CNN. This means that the classification accuracy using a CNN and only the gradient of acceleration for the foot-mounted sensor is as high as an OA of 75% and an F1-Score of 75%. In fact, the CNN outperforms RF in this case. More globally in this scenario, pressure signal channels (6 last variables of Fig. 4 and 5) were not important, neither were the integrals of acceleration and of angular rate. The order of importance of these channels also has shown variable tendencies depending on the placement of the sensors and the used algorithm.

2) *Best Model Selection:* According to the customized channel selection approach and to the channel by channel classification results, a first subset of features was selected. It was used with both RF and the CNN. As a result, all combination sets led to a decrease in the classification accuracy as compared with the gradient-acceleration CNN model. Therefore, the best model in the first scenario is the CNN based only on the gradient of acceleration, which requires only a 3-axis accelerometer and reduces the computational load. The detailed classification results using this model are provided in Table III that shows the mean and standard deviation (std) of the OA and F1-Score based on 5 test subjects. The model architecture and hyper-parameters are described below:

- 3 convolutional layers, with 30 filters each, and a respective kernel size of $3/8$ s for the first layer, $3/4$ s for the second, and then $3/2$ s for the last convolutional layer. This progressive increase in the kernel size was chosen to capture different patterns in the training samples. A maximum pooling of 2 was applied to each convolutional layer. The activation function relative to the convolutional layers was LeakyRelu with an Alpha (or slope) value tuned at 0.1.
- The 3 convolutional layers are followed by two dense layers of 200 and 50 neurons. Their activation function

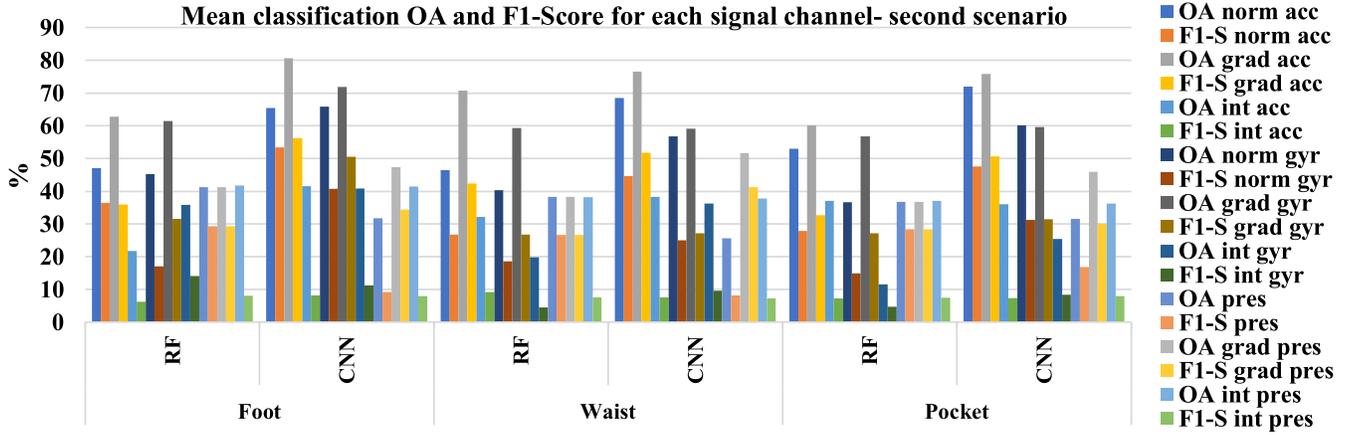


Fig. 5. Channel by channel classification for the second scenario.

was Relu. The output layer has as many neurons as the number of classes, depending on the considered scenario, and operates according to a softmax activation function.

- The cost function was the categorical cross-entropy. The optimizer was Adam, with an initial learning rate of 0.01.

C. Second Scenario Classification Results

1) *Channel by Channel Classification Results With RF and CNN*: For the second scenario, the channel by channel classification results are presented in Fig. 5. The first observation is that the F1-Score is systematically much lower than the OA as compared with the first scenario. This emphasizes the weak representation of the added classes in the second classification scenario, which are stairs and elevators. However, this lack of balance is quite relevant as it reflects the actual time distribution of the classes in daily activity. Indeed, time spent in public transport and while walking or biking is much longer than the time spent in stairs or elevators. For this reason, model selection is based on the F1-Score rather than OA values. This way, the time distribution of the classes is representative of a real scenario while model selection is not biased towards the major class.

The second observation is that, similar to the first scenario, the norm and gradient of acceleration, and the norm and gradient of angular rate are the most important channels. Yet, the difference between RF and CNN with respect to the pressure channels is considerable. In fact, RF ranks equally the F1-Score relative to the smoothed pressure and gradient of pressure, while CNN ranks the gradient of pressure above the smoothed pressure. The CNN also shows higher F1-Score values as compared to RF. In fact, most of the channels provide better results when they are used with the CNN. Therefore, the outcome is that the CNN performs better than RF on the same data, and second, the gradient of pressure seems to be more important as compared to the first scenario. According to Fig. 5, for all sensor placements, the F1-Score relative to the gradient of pressure is above 30%, while it was only about 10% in the first scenario for the foot and waist-attached sensors with the CNN model.

From this analysis, we decided to include the gradient of pressure with the previously selected signal channels, namely the norms and gradients of acceleration and angular rate.

TABLE III

RESULTS OF THE FIRST CLASSIFICATION SCENARIO: CNN WITH THE GRADIENT OF ACCELERATION

	Mean OA	std OA	Mean F1-Score	std F1-Score
Foot	75.63	3.24	75.57	3.53
Waist	69.14	9.75	67.05	10.93
Pocket	70.36	3.48	62.53	8.27

2) *Best Model Selection*: According to the channel by channel selection scheme, and from a first selection that includes the gradient of pressure to the norms and gradients of acceleration and of angular rate, the best results were obtained with the gradients of acceleration and of pressure. This result could be expected because of the importance of the gradient of pressure in detecting stairs and elevators as the signal directly relates to height change. In fact, the higher the slope, the faster the elevation changes. The selected model classification results are presented in Table IV. The best classification result is again obtained for the foot-mounted sensor with an OA of 79.41% and an F1-Score of 69.61%. The foot is additionally the placement that provides the F1-Score with the lowest standard deviation (std), meaning that the classification accuracy is close to the mean value for all test subjects, which means that the model is stable and robust.

D. Randomized Cross-Validation With CNN and RF

Results obtained with all subjects using the randomized cross validation method, with all sensor placements and in both classification scenarios, have shown high classification accuracy rates. As a reminder, the randomized cross-validation approach consists in merging data from all subjects and splitting it into two groups. 70% of these data are used for training, and 30% are used for testing. For the RF model, the latter were comprised between 89% and 99%, which is also the case for their corresponding F1-Scores (see Table V). For the CNN, the classification results are globally lower than those of RF. The highest accuracy was obtained with the sensor attached to the foot with an F1-Score of 90.98% in the first scenario and 90.42% in the second scenario. At this stage, it is important to notice the supremacy of RF over CNN, which is compliant with the state-of-the-art results with

TABLE IV

RESULTS OF THE SECOND CLASSIFICATION SCENARIO: CNN WITH THE GRADIENTS OF ACCELERATION AND OF PRESSURE

	Mean OA	std OA	Mean F1-Score	std F1-Score
Foot	79.41	7.07	69.61	9.34
Waist	76.63	6.98	64.41	12.57
Pocket	72.71	7.24	63.32	11.56

TABLE V

RESULTS OBTAINED USING THE RANDOMIZED CROSS-VALIDATION APPROACH WITH RF AND CNN

		Foot	Pocket	Waist
Scenario.1 - RF	OA	99.77%	99.66%	99.69%
	F1-Score	98.84%	99.21%	96.70%
Scenario.1 - CNN	OA	94.62%	95.17%	84.62%
	F1-Score	90.98%	89.84%	70.02%
Scenario 2. - RF	OA	99.54%	99.07%	98.80%
	F1-Score	96.94%	94.28%	93.98%
Scenario.2 - CNN	OA	98.48%	96.77%	96.60%
	F1-Score	90.42%	84.08%	85.83%

respect to both the evaluation method and classification performance. However, as shown earlier, these results are biased as the test subjects are not isolated from the train subjects. Therefore, these results should be put into the perspective of the randomized-cross validation method, which systematically overestimates the accuracy with respect to a real-life evaluation scenario.

VII. DISCUSSION

The above sections provided the results of 2 classifiers, RF and CNN, 2 classification scenarios, 3 sensor placements, 5 test subjects, 2 different feature sets, and 2 evaluation methods. Scenario 1 corresponds to: Still - Walk - Bike - Tramway - Bus. Scenario 2 corresponds to: Still - Walk - Bike - Elevator Up - Elevator down - Up Stairs - Downstairs - Tramway - Bus. According to the randomized cross-validation method, for the RF model, the OA was above 98% in all scenarios and for sensor placements. The F1-Score was comprised between 93.98% and 99.21%. For the CNN model, the OA ranged from 84.62% and 98.48%, all scenarios and all sensor placements included. In terms of F1-Score, they ranged from 70.02% to 90.98%. Given this evaluation method, the results are at least as high as those that can be found in the state-of-the-art for RF. The CNN model shows lower accuracy rates as compared with RF, which is compliant with the literature review. However, these results are biased as demonstrated in this study.

When test subjects are separated from train subjects, the best model, selected on the basis of the customized channel by channel selection method, was the CNN. In the first classification scenario, the highest average classification OA was 75.63% and the average F1-Score was 75.57%. These scores were obtained for the sensor being foot-mounted. For the second scenario, the highest average OA was 79.41% and the average F1-Score was 69.61% and they were also obtained for the sensor placed on the foot. According to the F1-Score values, the second optimal sensor placement is the waist, which shows an OA of 69.14% and an F1-Score of 67.05% in the first scenario and an OA of 76.63% and

an F1-Score of 64.41% in the second scenario. From these results, we conclude that in both scenarios, the best results are obtained with the IMU being either foot-mounted or attached to the waist. In fact, the foot-mounted sensor reflects the movement of the foot and is more suitable to distinguish body-induced activities such as walking, biking, or using stairs. Meanwhile, the waist is closer to the body center of mass and reflects the movement of the whole body instead of specific body limbs. More globally, the foot-mounted sensor has shown better results than the waist-attached sensor in both scenarios, which makes it the best placement for TMD purposes. Besides, we observe that the difference between randomized cross-validation and separate test-subject-based validation is huge with particularly overestimated accuracies in the first case, especially using RF. The results additionally demonstrate that feature selection significantly improves the classification results, and this is valid for all sensor placements and all scenarios. The selection through several channel by channel cross validation operations proved to be more efficient than PCA, MI, and KT.

Overall, the highest errors were obviously due to one specific confusion. In fact, being still was often confused with being inside a vehicle, which is rather likely to occur because people generally stay still inside vehicles. Yet, pressure could be relevant to distinguish between vehicle and static modes due to the speed of motion inside vehicles that can be observed in faster height changes and therefore in faster variations of pressure [35]. However, since the segmenting window size is quite short, especially if the terrain is flat, pressure variation is not always high enough to remove these confusions. In some studies, being still was first discarded before performing the classification of transportation modes such as in [20] where a threshold-based detection of static periods was performed. However, thresholding has its own limitations and being inside vehicles may also verify the threshold condition applied to acceleration. Besides, these results demonstrate that higher classifications accuracy is obtained if static phases are not considered together with other transportation modes [29], [33].

VIII. CONCLUSION

In this work, we have realized the classification of 7 transportation modes using a new set of sensors: a 3-axis accelerometer, a 3-axis gyroscope and a barometer. An important contribution was to construct and share an improved dataset so that the work can be reproduced and benchmarked. Second, we demonstrated that testing TMD models on randomly chosen samples without separating the test subjects from train subjects inevitably leads to overestimated accuracy. And finally, we have proposed two robust Adam-gradient-based CNN models that can predict accurately transportation modes with the second scenario including indoor activities. Future work will consist in introducing new emerging transportation modes in urban spaces and study their global impact on the classification accuracy.

IX. ACKNOWLEDGMENT

The sponsors had no involvement in the design of the study, the collection, analysis and interpretation of data, and

in writing the manuscript. This work further forms part of a broader translational and interdisciplinary research program, GaitAlps. The graphical content used in the abstract was designed by macrovector / Freepik and Flaticon.com.

REFERENCES

- [1] G. Xiao, Z. Juan, and C. Zhang, "Travel mode detection based on GPS track data and Bayesian networks," *Comput., Environ. Urban Syst.*, vol. 54, pp. 14–22, Nov. 2015.
- [2] H. Huang, Y. Cheng, and R. Weibel, "Transport mode detection based on mobile phone network data: A systematic review," *Transp. Res. C, Emerg. Technol.*, vol. 101, pp. 297–312, Apr. 2019.
- [3] H. Gong, C. Chen, E. Bialostozky, and C. T. Lawson, "A GPS/GIS method for travel mode detection in New York City," *Comput., Environ. Urban Syst.*, vol. 36, no. 2, pp. 131–139, Mar. 2012.
- [4] M. Nikolic and M. Bierlaire, "Review of transportation mode detection approaches based on smartphone data," in *Proc. Swiss Transp. Res. Conf.*, Monte Verità, Switzerland, May 2017, pp. 1–20.
- [5] I. Semanjski, S. Gautama, R. Ahas, and F. Witlox, "Spatial context mining approach for transport mode recognition from mobile sensed big data," *Comput., Environ. Urban Syst.*, vol. 66, pp. 38–52, Nov. 2017.
- [6] V. C. Coroamă, C. Türk, and F. Mattern, "Exploring the usefulness of Bluetooth and WiFi proximity for transportation mode recognition," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput., ACM Int. Symp. Wearable Comput.*, London, U.K., Sep. 2019, pp. 37–40.
- [7] M. C. Yu, T. Yu, S. C. Wang, C. J. Lin, and E. Y. Chang, "Big data small footprint: The design of a low-power classifier for detecting transportation modes," in *Proc. Int. Conf. Very Large Data Bases.*, Hangzhou, China, Sep. 2014, vol. 7, no. 13, pp. 1429–1440.
- [8] X. Liang and G. Wang, "A convolutional neural network for transportation mode detection based on smartphone platform," in *Proc. IEEE Int. Conf. Mobile Ad Hoc Sensor Syst.*, Orlando, FL, USA, Oct. 2017, pp. 338–342.
- [9] H. Zhao, C. Hou, H. Alrobassy, and X. Zeng, "Recognition of transportation state by smartphone sensors using deep bi-LSTM neural network," *J. Comput. Netw. Commun.*, vol. 2019, pp. 1–11, Jan. 2019.
- [10] X. Su, Y. Yao, Q. He, J. Lu, and H. Tong, "Personalized travel mode detection with smartphone sensors," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Seattle, WA, USA, Dec. 2017, pp. 1341–1348.
- [11] A. Vassilev, "Data mining applied to transportation mode classification problem," in *Proc. Int. Conf. Vehicle Technol. Intell. Transp. Syst.*, Funchal, Portugal, Mar. 2018, pp. 36–46.
- [12] C. Carpineti, V. Lomonaco, L. Bedogni, M. Di Felice, and L. Bononi, "Custom dual transportation mode detection by smartphone devices exploiting sensor diversity," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. Workshops*, Athens, Greece, Mar. 2018, pp. 367–372.
- [13] L. Wang and D. Roggen, "Sound-based transportation mode recognition with smartphones," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brighton, U.K., May 2019, pp. 930–934.
- [14] M. A. Shafique and E. Hato, "Use of acceleration data for transportation mode prediction," *Transportation*, vol. 42, no. 1, pp. 163–188, Jan. 2015.
- [15] D. Shin *et al.*, "Urban sensing: Using smartphones for transportation mode classification," *Comput., Environ. Urban Syst.*, vol. 53, pp. 76–86, Sep. 2015.
- [16] Y.-J. Byon, J. Ha, C.-S. Cho, T.-Y. Kim, and C. Yeun, "Real-time transportation mode identification using artificial neural networks enhanced with mode availability layers: A case study in Dubai," *Appl. Sci.*, vol. 7, no. 9, p. 923, Sep. 2017.
- [17] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Using mobile phones to determine transportation modes," *ACM Trans. Sensor Netw.*, vol. 6, no. 2, pp. 1–27, Feb. 2010.
- [18] A. Jahangiri and H. Rakha, "Developing a support vector machine (SVM) classifier for transportation mode identification using mobile phone sensor data," in *Proc. Transp. Res. Board 93rd Annu. Meeting*, Washington, DC, USA, Jan. 2014, pp. 1–14.
- [19] P. Widhalm, P. Nitsche, and N. Brandie, "Transport mode detection with realistic Smartphone sensor data," in *Proc. Int. Conf. Pattern Recognit.*, Tsukuba, Japan, Nov. 2012, pp. 573–576.
- [20] O. Lorintiu and A. Vassilev, "Transportation mode recognition based on smartphone embedded sensors for carbon footprint estimation," in *Proc. IEEE Conf. Intell. Transp. Syst.*, Rio de Janeiro, Brazil, Nov. 2016, pp. 1976–1981.
- [21] L. Bedogni, M. Di Felice, and L. Bononi, "Context-aware Android applications through transportation mode detection techniques," *Wireless Commun. Mobile Comput.*, vol. 16, no. 16, pp. 2523–2541, 2016.
- [22] S.-H. Fang, Y.-X. Fei, Z. Xu, and Y. Tsao, "Learning transportation modes from smartphone sensors based on deep neural network," *IEEE Sensors J.*, vol. 17, no. 18, pp. 6111–6118, Sep. 2017.
- [23] S.-H. Fang *et al.*, "Transportation modes classification using sensors on smartphones," *Sensors*, vol. 16, no. 8, p. 1324, Aug. 2016.
- [24] L. Wang, H. Gjoreski, M. Ciliberto, S. Mekki, S. Valentin, and D. Roggen, "Enabling reproducible research in sensor-based transportation mode recognition with the Sussex-Huawei dataset," *IEEE Access*, vol. 7, pp. 10870–10891, 2019.
- [25] L. Wang *et al.*, "Summary of the Sussex-Huawei locomotion-transportation recognition challenge 2019," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput. Int. Symp. Wearable Comput.*, London, U.K., Sep. 2019, pp. 849–856.
- [26] M. Ahmed, A. D. Antar, T. Hossain, S. Inoue, and M. A. R. Ahad, "POIDEN: Position and orientation independent deep ensemble network for the classification of locomotion and transportation modes," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput. Int. Symp. Wearable Comput.*, London, U.K., Sep. 2019, pp. 674–679.
- [27] G. Ascì and M. A. Guvensan, "A novel input set for LSTM-based transport mode detection," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. Workshops (PerCom Workshops)*, Kyoto, Japan, Mar. 2019, pp. 107–112.
- [28] X. Liang, Y. Zhang, G. Wang, and S. Xu, "A deep learning model for transportation mode detection based on smartphone sensing data," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 12, pp. 5223–5235, Nov. 2019.
- [29] H. I. Ashqar, M. H. Almannaa, M. Elhenawy, H. A. Rakha, and L. House, "Smartphone transportation mode recognition using a hierarchical machine learning classifier and pooled features from time and frequency domains," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 1, pp. 244–252, Jan. 2019.
- [30] A. Jahangiri and H. A. Rakha, "Applying machine learning techniques to transportation mode recognition using mobile phone sensor data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2406–2417, Oct. 2015.
- [31] H. Gjoreski *et al.*, "The university of Sussex-Huawei locomotion and transportation dataset for multimodal analytics with mobile devices," *IEEE Access*, vol. 6, pp. 42592–42604, 2018.
- [32] H. Wang, H. Luo, F. Zhao, Y. Qin, Z. Zhao, and Y. Chen, "Detecting transportation modes with low-power-consumption sensors using recurrent neural network," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov.*, Guangzhou, China, Dec. 2018, pp. 1098–1105.
- [33] S. Ballı and E. A. Sağbaşı, "Diagnosis of transportation modes on mobile phone using logistic regression classification," *IET Softw.*, vol. 12, no. 2, pp. 142–151, Apr. 2018.
- [34] M. Hossain and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Mining Knowl. Manage. Process.*, vol. 5, no. 2, pp. 1–11, 2015.
- [35] A. Manivannan, W. Chin, A. Barrat, and R. Bouffanais, "On the challenges and potential of using barometric sensors to track human activity," *Sensors*, vol. 20, no. 23, pp. 1–28, Nov. 2020.