Combining LSTM and CNN for Mode of Transportation Classification from Smartphone Sensors

Björn Friedrich*
bjoern.friedrich@uni-oldenburg.de
Carl von Ossietzky University
Oldenburg, Germany

Carolin Lübbe*
carolin.luebbe@uni-oldenburg.de
Carl von Ossietzky University
Oldenburg, Germany

Andreas Hein
andreas.hein@uni-oldenburg.de
Carl von Ossietzky University
Oldenburg, Germany

ABSTRACT

The broad availability of smartphones and Inertial Measurement Units in particular brings them into focus of recent research. Inertial Measurement Unit data is used for a variety of tasks. One important task is the classification of the mode of transportation. In this paper, we present a deep-learning-based algorithm, that combines long-short-term-memory (LSTM) layer and convolutional layer to classify eight different modes of transportation on the Sussex-Huawei Locomotion-Transportation (SHL) dataset. The inputs of our model are the accelerometer, gyroscope, linear acceleration, magnetometer, gravity and pressure values as well as the orientation information. We achieve a F_1 score of 98.96 % on our private test set. We participated as team $103114102106|_8$ in the Sussex-Huawei Locomotion-Transportation (SHL) recognition challenge.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing**; • **Computing methodologies → Neural networks**.

KEYWORDS

SHL dataset, convolutional neural networks, supervised learning, IMU, mode of transportation, classification

*Both authors contributed equally to this research.

UbiComp/ISWC '20 Adjunct, September 12–16, 2020, Virtual Event, Mexico

ACM ISBN 978-1-4503-8076-8/20/09…$15.00
https://doi.org/10.1145/3410530.3414350

ACM Reference Format:
Björn Friedrich, Carolin Lübbe, and Andreas Hein. 2020. Combining LSTM and CNN for Mode of Transportation Classification from Smartphone Sensors. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers (UbiComp/ISWC '20 Adjunct), September 12–16, 2020, Virtual Event, Mexico.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3410530.3414350

1 INTRODUCTION

The broad acceptance of smartphones holds the potential for large scale human centered sensing and research. Most smartphones are capable of positioning themselves in a global frame of reference, e.g. GPS, but the accuracy depends on the signal quality and line of site between the sensor and the satellites. The accuracy decreases significantly indoors or underground, as well as the features derived from the measurements. Inertial Measurement Units (IMU) are not reliant on external infrastructure. On the one hand, the data quality of the IMU does not depend on whether the sensor is underground or not, and on the other hand, the IMU data depend on the kinematic chain between the sensor and the source of the force applied to the sensor.

The task of the SHL recognition challenge 2020, which is is being organised for the third time [11, 14], is to classify the mode of transportation using IMU data from smartphone sensors. For this purpose, a classifier must be built that it works reliably independent of the location of the smartphone on a person's body. The training and validation sets provided comprise data from four smartphones carried at different locations on the body, and the test set consists of data from one smartphone worn on one location. The location and mode of transportation labels are unknown.

The structure of the paper is as follows: the next section gives a brief overview of the state of the art. In Section 3 the used dataset, the pre-processing pipeline, the used algorithm and the used computational resources are described. Thereafter the results are shown and discussed in the subsequent section. At the end we draw conclusions and giving some prospects for future work.

2 STATE OF THE ART

For several years, extensive work on understanding and sensing the mobility behaviour of people has been carried out. This section introduces the state of the art related to mode of transportation classification using machine learning approaches based on smartphone sensor data. All approaches that include contextual information are not considered, since this research focuses on the use of information derived from smartphone sensors. A common approach is to understand the detection of the mode of transportation as a classification problem. We have assigned related works to the following two categories: 1. traditional machine learning-based classification and 2. deep learning-based classification.

Different traditional machine learning algorithms are used to classify the mode of transportation. Antar et al. [1] introduced a random forest (RF) classification algorithm to detect transportation mode on the SHL dataset and obtained an accuracy of 92 %. Likewise, Liono et al. [7] employed RFs to differentiate six transportation modes (bus, light rail, car, scooter, escalator, elevator) using the phone's screen status, an accelerometer, magnetic field and light sensors. On the Crowdsignals dataset they achieved an accuracy of 91 %. Fang et al. [2] proposed 14 different handcrafted features and used three machine learning algorithms including decision trees (DT), k-nearest neighbour (kNN) and support vector machine (SVM) to classify the user's transportation and vehicular modes. For this purpose an accelerometer, magnetometer and a gyroscope were used. In the transportation mode classification, SVM shows the best performance in accuracy with 86 % on five classes (vehicle, bike, run, still, walk). Hemminki et al. [5] used a traditional approach that aims to identify five transportation modes (bus, train, metro, tram and car) and achieved an accuracy of 84 %.

Recently, large scale datasets became available and enabled the application of deep learning techniques. The deep learning algorithms are outperforming the traditional approaches using handcrafted features. Jeyakumar et al. [6] proposed a deep convolutional bidirectional-LSTM ensemble trained directly on raw sensor data on the SHL dataset. Using this approach, an F1-score of 96 % was achieved for transportation mode classification. Qin et al. [8] introduced a deep-learning-based algorithm that combines a CNN and LSTM network. By using CNN-extracted and handcrafted features (i.e. segment and peak features), the algorithm is able to distinguish the transportation modes with an accuracy of 98.1 % on the SHL dataset. Vu et al. [10] proposed a gate-based recurrent neural network to detect the transportation mode on the HTC dataset. This accelerometer-based approach achieved an accuracy of 94.72 %. Tambi et al. [9] presented a CNN that distinguishes four transportation modes (bus, car, subway, train) by using mobile sensor data derived from an accelerometer and gyroscope in the spectral domain. An accuracy of 91 % was achieved. Using an LSTM network for time series classification Friedrich et al. [3] achieved an F_1 score of 65 % on the SHL dataset.

3 MATERIALS AND METHODS

Dataset

Part of the Sussex-Huawei Locomotion- Transportation (SHL) dataset [4, 13] provided contains data from smartphones carried on the body in various positions. The dataset was collected with three participants over 31.6 *d*, each of them carrying four phones positioned at four locations *hand*, *bag*, *hips*, and *torso*. The values of the hardware sensors *accelerometer*, *gyroscope*, *magnetometer*, and *pressure* as well as the software sensor values of *linear acceleration*, *gravity* and *orientation*. The measurement frequency was 100 *Hz*. Each individual sensor value was labeled, i.e. 100 labels are available for 1*s*. The dataset includes eight different modes of transportation *still*, *walk*, *run* ,*bike* ,*car*, *bus*, *train*, and *subway*. The samples are consecutive in time for the training and validation set, as opposed to the test set. The training data comprises the values of all four phone locations from one participant and the validation data the values of the other two participants from all locations as well. The test set contains data from the users of the validation set, but only from one unknown phone location. Overall there are 196.072 training samples, 28.789 validation samples and 57.573 test samples. Moreover, the dataset has a large class imbalance.

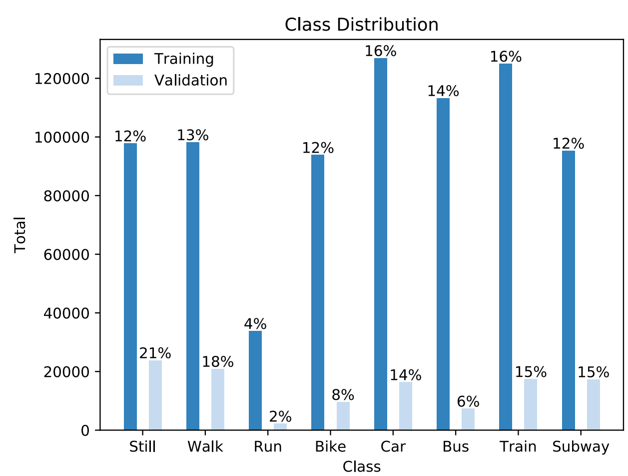

Figure 1: The histogram of the distribution of the labels in the training and validation set.

Pre-processing

Before pre-processing we performed some data integrity checks. We found that the labels for some samples are not uniform, i.e. the samples contain transitions of modes of transportation. Since the amount of samples containing a transition was less than 1 % we assigned the label by majority decision. Thus, our dataset has only one label instead of 500 for each sample. Then, the training set has been merged with the validation set. To overcome the class imbalance we followed a simple approach and oversampled by copying random samples and undersampled by deleting random samples. We used 30.000 samples, because the number of classes in which samples had to be deleted equals the number of classes in which samples had to be copied. After balancing, the full dataset was split into new training, validation and test sets in a stratified way. 75 % of the full dataset was assigned to the training set, 15 % to the validation set and the remaining 10 % to the test set. Finally, the data from all phone locations were merged. The training set contains 720.000 samples, the validation set 144.000 samples and the test set 96.000 samples.

Two pre-processing steps are applied on the balanced dataset. The first one is to apply a low-pass filter on all data. We use a second order filter with a cut-off frequency of 25 *Hz*. The second step is standard scaling by subtracting the mean and by dividing by the variance. Standard scaling is applied to each feature in each dimension separately.

Algorithm

The architecture we propose, see Figure 2, combines an augmentation and an LSTM layer as well as several convolutional and fully-connected layers to perform transportation mode classification. The input data is split into seven streams, one stream per sensor. To artificially increase the number of training samples, an augmentation layer is implemented, which augments four windows of size 50 of each sample with a factor of 2. This is followed by an LSTM layer, that is able to store information about time to find temporal correlations of the input sequences. The LSTM layer comprises 64 neurons, sigmoid recurrent activation and tanh activation. It is followed by a dropout layer, with a dropout rate of 0.25, that is used to avoid overfitting, a convolution layer and at the end of each stream a maximum pooling layer. The convolutional layer consists of 128 filters, a kernel size 8, stride length 2 and a Leaky ReLU activation function with $\alpha = 0.001$. Maximum pooling was performed with stride length 2. The seven streams are then merged via a concatenation layer, which allows us to combine all features to extract meaningful information. Afterwards, a convolutional layer and a max pooling layer are used 4 times in a row, whereupon a flatten layer completes the second block (see Figure 3). In all type 2 blocks, maximum pooling, the convolutional stride and the Leaky ReLu activation with $\alpha = 0.001$ were the same. The number of filters and the filter size were arranged in ascending order 64, 64, 128, 128 and 16, 32, 64, 64. The subsequent fully connected layers, each followed by a dropout layer, recombine the representations learned from the convolution layer and reduce the dimension. Both blocks of type 3 used the same parameters. The dense layer had 256 neurons, the dropout rate was 0.25 and Leaky ReLU was used as activation function, as before. In the last step, the classification layer uses the softmax activation function for the mode of transport classification.

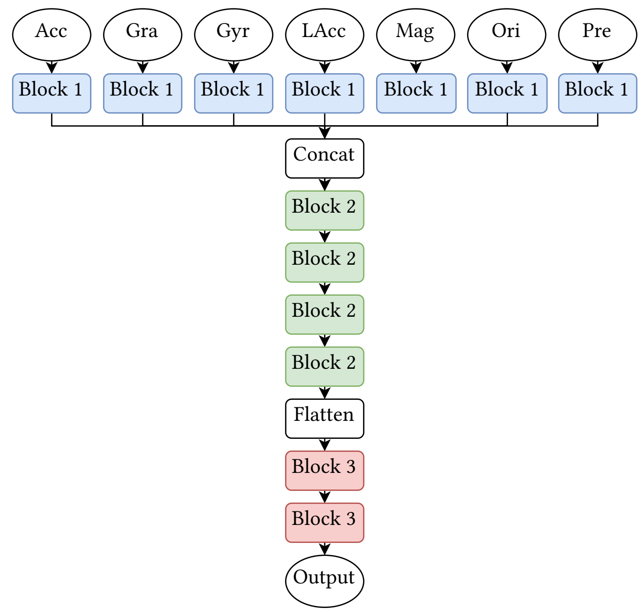

Figure 2: The architecture of the model. Each sensor modality had its own input and the intermediate features were fused in the concatenation layer in the second dimension.

Computational Resources

Our computational resources were comprised of two mobile computers and a high performance computing (HPC) cluster. Debugging the network architectures with a small number of samples were performed on the mobile computers. We have had a node of the of the HPC cluster exclusively reserved for our experiments and the final training. The mass storage of the reserved node is 1 TB. An overview of the key components of our hardware is shown in Table 1.

4 RESULTS

The final training took 3 *d* and 13 *h* for 100 epochs. We used categorical crossentropy loss and the F_1 score as metric. The Adam optimisation algorithm was used for gradient optimisation and we used a learning rate schedule with exponential

Table 1: Hardware Overview

Part	Lenovo NeXtScale System	Dell XPS 15	Dell Latitude 5401
CPU	2x Intel Xeon E5-2650 12x2.2GHz	Intel Core i9 8x2.4GHz	Intel Core i7 6x2.6GHz
GPU 1	NVIDIA Tesla P100 16GB HBM2	Intel UHD Graphics 630	Intel UHD Graphics 630
GPU 2	*not available*	NVIDIA Geforce GTX 1650 4GB GDDR6	*not available*
GPU 3	*not available*	NVIDIA TITAN V 12GB HBM2 (Thunderbolt 3)	*not available*
RAM	256GB DDR4 2400MHz	32GB DDR4 2677MHz	16GB DDR4 2677MHz
Mass Storage	1TB HDD	1TB NVMe SSD	512GB NVMe SSD

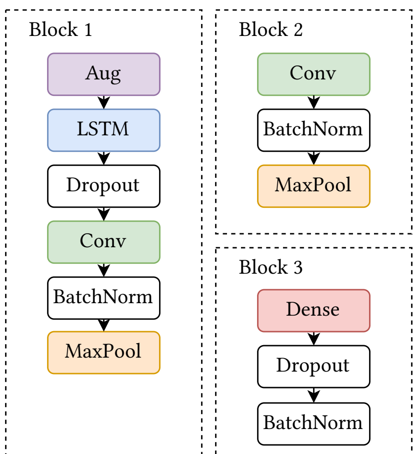

Figure 3: A detailed view of the three different blocks of layers used in our architecture.

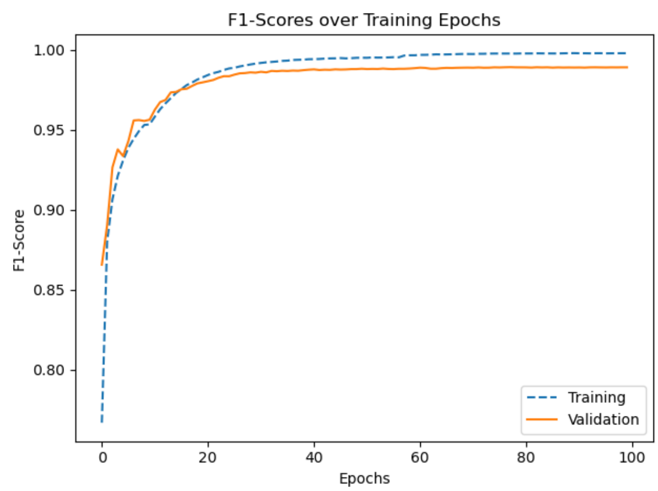

Figure 4: The progress of the score for the final training for 100 epochs. The progress shows an asymptotic behaviour after around about 40 epochs.

decay after the first 10 epochs with an an initial learning rate of 0.001. After some preliminary experiments we found that the model has difficulties with distinguishing between the classes *train* and *subway*. So, we put a higher weight (3x) on the gradient update for the class *train*. One epoch with a batch size of 500 samples per batch takes about 52 *min* and the prediction on the challenge test set 146.18 *s* on the Lenovo NeXtScale System described in Table 1. The Figures 4 and 5 show the graphs of the F_1 score and the loss of the final training. In the beginning the score and the loss have a high slope and later on the slope is asymptotically approaching the limit 0. During the first 10 epochs the validation score is slightly better than the training score and the validation loss is slightly smaller than the training loss. The confusion matrix shows that the model performs best on the classes *walk* and *run* and worst on the classes *still* and *subway*.

The best epoch was epoch 77 with a validation score of 98.93 % and a score of 98.96 % on our private test set.

5 DISCUSSION

The progress' of the training score and the training loss and the validation score and the validation loss are as expected and does not have any strong fluctuations. The reason for misclassifying *subway* as *train* might be due similar characteristics of both modes of transportation. It is possible that samples of the class *still* misclassified, because of the noise and bias of the sensors. The characteristic of no motion is the absence of any force except for gravity. The noise and the bias could superimpose the characteristic and hence lead to misclassification.

6 CONCLUSION & FUTURE WORK

This contribution introduced a machine learning approach for classifying eight different modes of transportation using

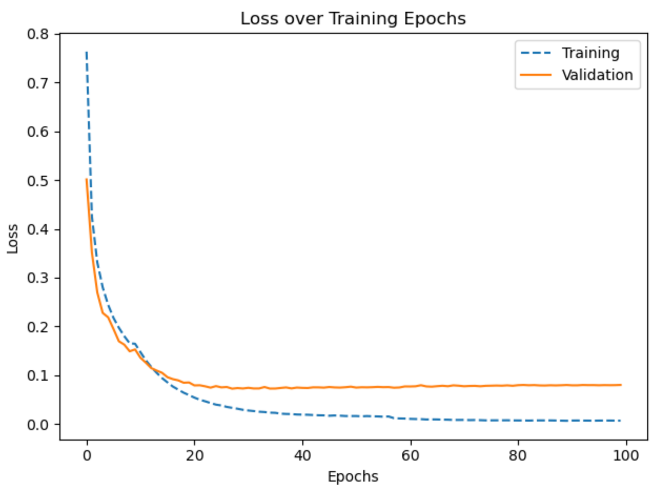

Figure 5: The progress of the loss for the final training for 100 epochs. The progress shows an asymptotic behaviour after around about 40 epochs. The progress corresponds to the progress of the score.

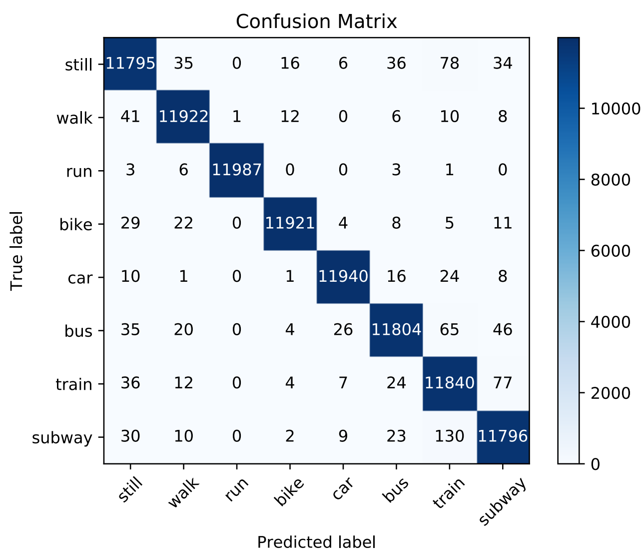

Figure 6: The Confusion Matrix for the internal test set. The classes with the most false classifications *still* and *subway*. The classes with the best true classifications are *run* and *walk*.

smarthpone IMU sensor data in the SHL Challenge 2020. We used the data of seven sensors, namely *accelerometer*, *gyroscope*, *magnetometer*, *linear acceleration*, *gravity*, *orientation* and *pressure*. Our model was combined of LSTM and convolutional layers and we introduced a layer for time series augmentation during runtime. The model achieve a F_1 score of 98.96 % on our test set.

To improve the classification score the characteristics of the sensors could be analysed. Using the samples of the class *still* may give useful information about the noise of the accelerometers and the drift of the gyroscopes of the smartphones. The information can be used for advanced pre-processing.

The recognition result for the testing dataset will be presented in the summary paper of the challenge [12].

ACKNOWLEDGMENTS

The final experiments were performed at the HPC Cluster CARL, located at the University of Oldenburg (Germany) and funded by the DFG through its Major Research Instrumentation Programme (INST 184/157-1 FUGG) and the Ministry of Science and Culture (MWK) of the Lower Saxony State. Moreover, we gratefully acknowledge the support of the NVIDIA Corporation with the donation of the TITAN V GPU used for this research.

REFERENCES

[1] Anindya Das Antar, Masud Ahmed, Mohammad Shadman Ishrak, and Md. Atiqur Rahman Ahad. 2018. A Comparative Approach to Classification of Locomotion and Transportation Modes Using Smartphone Sensor Data. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (Singapore, Singapore) *(UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 1497–1502. https://doi.org/10.1145/3267305.3267516

[2] Shih-Hau Fang, Hao-Hsiang Liao, Yu-Xiang Fei, Kai-Hsiang Chen, Jen-Wei Huang, Yu-Ding Lu, and Yu Tsao. 2019. Transportation Modes Classification Using Sensors on Smartphones,. *Sensors* 8 (19 August 2019), 1324–1339. Issue 16. https://doi.org/10.3390/s16081324

[3] Björn Friedrich, Benjamin Cauchi, Andreas Hein, and Sebastian Fudickar. 2019. Transportation Mode Classification from Smartphone Sensors via a Long-Short-Term-Memory Network. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) *(UbiComp/ISWC '19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 709–713. https://doi.org/10.1145/3341162.3344855

[4] H. Gjoreski, M. Ciliberto, L. Wang, F.J.O. Morales, S. Mekki, S. Valentin, and D. Roggen. 2018. The University of Sussex-Huawei locomotion and transportation dataset for multimodal analytics with mobile devices,. *IEEE Access* 6 (23 July 2018), 42592–42604.

[5] Samuli Hemminki, Petteri Nurmi, and Sasu Tarkoma. 2013. Accelerometer-Based Transportation Mode Detection on Smartphones. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems* (Roma, Italy) *(SenSys '13)*. Association for Computing Machinery, New York, NY, USA, Article 13, 14 pages. https://doi.org/10.1145/2517351.2517367

[6] Jeya Vikranth Jeyakumar, Eun Sun Lee, Zhengxu Xia, Sandeep Singh Sandha, Nathan Tausik, and Mani Srivastava. 2018. Deep Convolutional Bidirectional LSTM Based Transportation Mode Recognition. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (Singapore, Singapore) *(UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 1606–1615. https://doi.org/10.1145/3267305.3267529

[7] Jonathan Liono, Zahraa S. Abdallah, A. K. Qin, and Flora D. Salim. 2018. Inferring Transportation Mode and Human Activity from Mobile Sensing in Daily Life. In *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services* (New York, NY, USA) *(MobiQuitous '18)*. Association for Computing Machinery, New York, NY, USA, 342–351. https://doi.org/10.1145/3286978.3287006

[8] Y. Qin, H. Luo, F. Zhao, C. Wang, J. Wang, and Y. Zhang. 2019. Toward Transportation Mode Recognition Using Deep Convolutional and Long Short-Term Memory Recurrent Neural Networks. *IEEE Access* 7 (2019), 142353–142367.

[9] Ritiz Tambi, Paul Li, and Jun Yang. 2018. An Efficient CNN Model for Transportation Mode Sensing. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems* (Shenzhen, China) *(SenSys '18)*. Association for Computing Machinery, New York, NY, USA, 315–316. https://doi.org/10.1145/3274783.3275160

[10] Toan H. Vu, Le Dung, and Jia-Ching Wang. 2016. Transportation Mode Detection on Mobile Devices Using Recurrent Nets. In *Proceedings of the 24th ACM International Conference on Multimedia* (Amsterdam, The Netherlands) *(MM '16)*. Association for Computing Machinery, New York, NY, USA, 392–396. https://doi.org/10.1145/2964284.2967249

[11] Lin Wang, Hristijan Gjoreski, Mathias Ciliberto, Paula Lago, Kazuya Murao, Tsuyoshi Okita, and Daniel Roggen. 2019. Summary of the Sussex-Huawei Locomotion-Transportation Recognition Challenge 2019. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) *(UbiComp/ISWC '19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 849–856. https://doi.org/10.1145/3341162.3344872

[12] L. Wang, H. Gjoreski, M. Ciliberto, P. Lago, K. Murao, T. Okita, and D. Roggen. 2020. Summary of the Sussex-Huawei Locomotion-Transportation Recognition Challenge 2020. In *Proceedings of the 2020 ACM International Joint Conference and 2020 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp '20)*.

[13] L. Wang, H. Gjoreski, M. Ciliberto, S. Mekki, S. Valentin, and D. Roggen. 2019. Enabling reproducible research in sensor-based transportation mode recognition with the Sussex-Huawei dataset,. *IEEE Access* 7 (03 January 2019), 10870–10891.

[14] Lin Wang, Hristijan Gjoreskia, Kazuya Murao, Tsuyoshi Okita, and Daniel Roggen. 2018. Summary of the Sussex-Huawei Locomotion-Transportation Recognition Challenge. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (Singapore, Singapore) *(UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 1521–1530. https://doi.org/10.1145/3267305.3267519