Ger J Exerc Sport Res 2022 · 52:248–252
https://doi.org/10.1007/s12662-022-00817-y
Received: 31 October 2021
Accepted: 13 April 2022
Published online: 6 May 2022

Eni Hysenllari · Jörg Ottenbacher · Darren McLennan
movisens GmbH, Karlsruhe, Germany

Validation of human activity recognition using a convolutional neural network on accelerometer and gyroscope data

Supplementary Information

The online version of this article (https://doi.org/10.1007/s12662-022-00817-y) contains supplementary material, which is available to authorized users.

Background

Systems that recognize human activities from sensors worn on the body open possible healthcare applications, such as eldercare support, fitness monitoring, chronic care, long-term preventive and cognitive assistance. The maintenance of physical activity proves highly beneficial for health throughout the lifespan (Buchner, 2014). The need to provide solutions for real-life applications such as healthcare and care of the elderly makes this a popular strategy (Rana, Shetty, & Jha, 2015).

One way to achieve human activity recognition (HAR) is by means of machine learning (ML). In a classical ML context, HAR uses hand-crafted features, which comprises information extracted from a signal, and bases the recognition on classification techniques like support vector machines (SVM). Extracting features requires a deep understanding of data, and it can be time-consuming and difficult to implement. The recent increase in big data allows the development of deep learning methodologies, such as convolution neural network. This application not only extracts features on its own, but also shows in many areas, like HAR and image recognition, greater success in recognition accuracy with increased performance and state-of-the-art results (Shakya, Zhang, & Zhou, 2018; Gadri & Neuhold, 2020).

We selected wearable sensors and thus sensor-based recognition from the available choices (e.g. cameras, wearables) for classifying and capturing activity (Zhou, Li, & He, 2015; Ji, Cheng, Feng, & Tao, 2018; Zhang, Yang, Lin, & Zhu, 2012). Generally, accelerometers provide suitable data for assessing the intensity of movements and extracting information on the overall quantity of physical activity (Staudenmayer, Zhu, & Catellier, 2012). We used raw acceleration and raw gyroscopic data as inputs for the convolutional neural network. The network extracts information from the inputs and uses the extracted information to predict the corresponding activity with a high certainty. The activities detected depend on the position of the sensor on the participant. Due to this, multiple Move 4 (movisens GmbH, Karlsruhe, Germany) sensors positioned on six different well-defined body parts acquire the data used to build the database.

Studies show that when using deep convolutional neural network (CNN) on acceleration and angular velocity data, accuracies achieved range from 95 to 97.62% (Cho & Yoon, 2018; Ronao & Cho, 2016; Jiang & Yin, 2015). Most studies (Rana et al., 2015; Avilés-Cruz, Ferreyra-Ramírez, Zúñiga-López, & Villegas-Cortéz, 2019; Shakya et al., 2018; Cho & Yoon, 2018; Ronao & Cho, 2016; Jiang & Yin, 2015) use ML techniques to recognize basic activities like walking, jogging, sitting, standing and lying but cycling is usually overlooked. In this article we present a CNN capable of recognizing walking, jogging, cycling, sitting, standing and lying activities by only using the acceleration and angular velocity signals obtained from one Move 4 sensor. Here we also focus on observing how the performance of neural network changes with sensor location. Our test data show that the accuracy of the CNN varies from 96.57% (ankle) to 99.28% (thigh).

Methods

Database

The database includes measurements from two studies. In the first study, 20 adults followed a structured study protocol while performing a series of 32 consecutive conditions (Giurgiu et al., 2020). The volunteers completed fully standardized activities and semistandardized activities. They performed each activity for 2 min in a laboratory setting or outdoor area. Further details and sample characteristics have been previously published (Giurgiu et al., 2020).

The database of the second study includes measurements from 15 participants, who followed a structured protocol while performing 20 different fully standardized and semistandardized activities in a laboratory setting or outdoor area. The volunteer group involved 9 men, varying in age from 22–60 years old with a body mass index (BMI) from 18.85–27.77 kg/m^2, and 6 women varying in age from 24–35 years old with a BMI

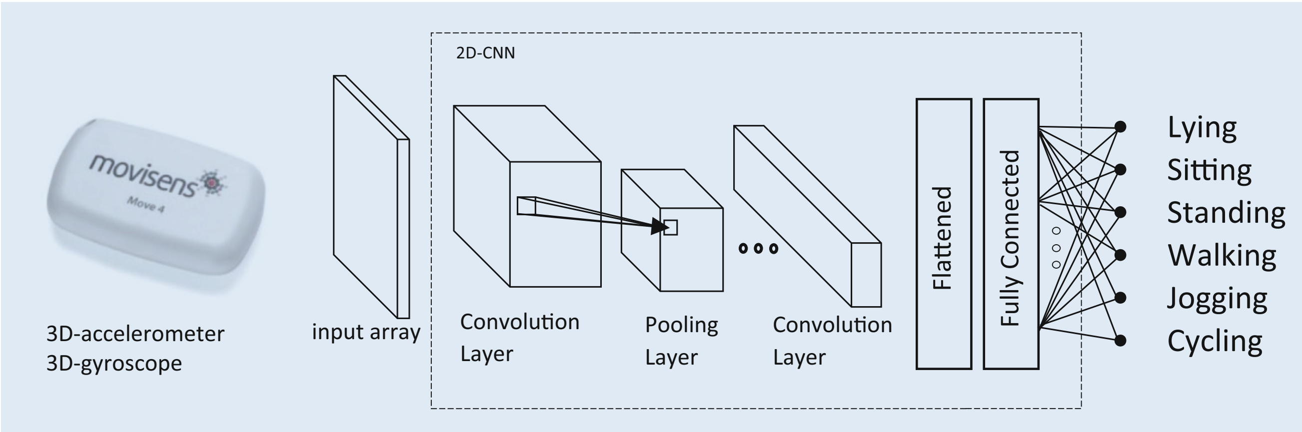

Fig. 1 ▲ The three-axis accelerometer Move 4 (movisens GmbH) captures the acceleration and angular velocity of the movement. After preprocessing, the data are arranged in an array that serves as input for the two-dimensional convolutional neural network. The convolution neural network makes an accurate prediction for every 4 s data stream

from 20.9–22.56 kg/m^2. Each participant performed each activity for 3–5 min.

We built the reference by mapping all conditions from both fully and semi-standardized sections to six basic activities—walking, jogging, cycling, sitting, standing, lying. For example, if the participant read the newspaper while sitting, then we labeled the duration of this condition as sitting.

Prior to commencing measurements in both studies, all monitors were placed correctly and their time clocks were synchronized. The sensors were worn on different locations on the body, including ankle, outer thigh, right side hip, chest, wrist and upper-arm. The Move 4 sensors, placed on six well-defined body parts, recorded the acceleration and angular velocity signals. The Move 4 is a triaxial activity monitor with a size of 62 × 39 × 11 mm (movisens GmbH, 2018). The sensor records acceleration within a range of ± 16 g and at a sampling rate of 64 Hz. It records angular velocity within a range of ± 2000 °/s and at a sampling rate of 64 Hz.

To train the CNN, we input the activity reference together with both the acceleration and angular velocity signals to allow it to make predictions on unseen data.

Preprocessing

Using convolutional neural networks removes the burden of extracting features by hand. The network performs this task independently and classifies the data based on the extracted features. In spite of this, some amount of preprocessing improves the performance of the CNN. This preprocessing phase helps to optimize the end results.

Depending on the participant, the sensors were placed on the right or left side of the body, meaning that the values of the x, y and z sensor axis do not represent the same directions for everyone. Thus, as a first step we calculated the acceleration in forward, right and down directions, and the angular velocity for leaning right, turning right and leaning backward rotation directions.

The Move 4 samples these two signals—three-dimensional (3D) acceleration and 3D angular velocity—at a frequency of 64 Hz. Studies show that this level of precision is not always required to achieve a high level of accuracy, citing a good prediction value even with sample rates as low as 16 Hz (Hamm, Stone, Belkin, & Dennis, 2012; Qi, Keally, Zhou, Li, & Ren, 2013). Down-sampling the signal to 16 Hz allows for improved functionality on the sensor (edge processing) without a loss of quality. After down-sampling the acceleration and gyroscope signal to 16 Hz, we normalized the data to have values varying from minus one to one. This helps with numerical issues and it makes it easier for the optimizer to find the optimal parameters.

Convolutional neural network

Before feeding the signals to the network for training, testing and later inference they have to be rearranged. Inference is the process of running live data points into an ML algorithm, in this case the 2D-CNN, to calculate an output such as a single numerical score that can represent a given class. We divided the database for training and testing purposes, by randomly choosing 14 participants from the first study and 11 participants from the second study to train the network. The data from the remaining participants were used to evaluate the already trained network. We found 4 s data segments provided a good compromise between accuracy and temporal resolution. Therefore, the rearrangement of input data involved dividing it into 4 s segments, with each segment describing one activity class and arranged in a way that helps the network extract relevant information. We made sure the network captured the dependency between acceleration and angular velocity by not merging these inputs as one vector but presenting them to the network separately. The network also analyses the local dependency over time, meaning it learns how the values of acceleration and angular velocity change in all three directions over the 4 s window for each activity class.

The steps we followed to recognize human activity are shown in ▫ **Fig. 1**, first feeding the generated input data to the network for training and testing. Here we

Abstract

Ger J Exerc Sport Res 2022 · 52:248–252 https://doi.org/10.1007/s12662-022-00817-y

E. Hysenllari · J. Ottenbacher · D. McLennan

Validation of human activity recognition using a convolutional neural network on accelerometer and gyroscope data

Abstract

Background. Human activity recognition (HAR) means identifying sequences of data recorded by specialized wearable sensors into known, well-defined classes of physical activity. In principle, activity recognition provides great societal benefits, especially in real-life, humancentric applications such as healthcare and care of the elderly. Using raw acceleration and angular velocity to train a convolutional neural network shows great success in recognition accuracy. This article presents the quality of activity recognition obtained using convolutional neural network on acceleration and angular velocity data recorded from different sensor locations.
Methods. Thirty-five volunteers from two studies (16 women and 19 men) with an average age of 28.54 years wore Move4/EcgMove4 accelerometers on 6 different body positions (ankle, thigh, hip, wrist, upper arm, chest) while completing typical activities (sitting, standing, lying, walking, jogging, cycling). We then used those databases to evaluate a two-dimensional convolutional neural network (2D-CNN) that takes 3D acceleration and 3D angular velocity signals as inputs to recognize human activity. We measure the networks performance using accuracy and Cohen's κ.
Results. Depending on the location of the sensor, the accuracy of the network varies from 96.57% (ankle) to 99.28% (thigh) and Cohen's κ varies from 0.96 (ankle) to 0.99 (thigh).
Conclusions. The performance of the 2D-CNN concerning human activity recognition showed excellent results. Using raw signals may enable real-time, on-device—also known as at the edge—activity recognition even in small devices with low computational power and small storage.

Keywords

Machine learning · Accelerometer · Raw acceleration · Angular velocity · Convolutional neural network

built a deep network rather than a wide one, since deeper models learn more effectively from training data, while a wider model falls prey to overfitting, effectively memorizing the data. A wide network performs well with data it has already seen, but struggles with new and novel inputs. Therefore, we increased the filter size of convolution layers gradually. By using different types of layers, we made sure that the network is robust towards noise and not prone to overfitting. After flattening the final output of the last convolution layer, we added a fully connected layer which determines the most likely activity class that occurred in the 4 s segment. The paper "An introduction to convolutional neural networks" by O'Shea and Nash describes how these layers work and their purpose (O'Shea & Nash, 2015).

The training process of a model is probably the most delicate part. The training dataset enables the CNN to learn the parameters while the testing dataset remains unseen and only used for testing and inference. We used the Adam approach to optimize the parameters and made sure the network learns slowly, by setting the initial rate to 0.001. We implemented the proposed CNN using Python programming language and the open source TensorFlow platform. Training a model requires a lot of computation power that is more suited to a graphics processing unit (GPU). In this case, we performed our calculations on the NVIDIA GeForce RTX 2080 graphic card (Nvidia, Santa Clara, CA, USA).

Statistics

We evaluate the CNN models using accuracy, sensitivity, specificity and Cohen's κ (CK). Accuracy refers to how close a measurement is to the true value. It is a ratio of correctly predicted observation to the total observations. Sensitivity and specificity describe the accuracy of a test which reports the presence or absence of a condition. Sensitivity measures how well a model identifies true positives and specificity measures how well a model identifies true negatives.

Another evaluation method we used is CK, which is a dimensionless index that can be used to express the agreement between two raters in a single number (Warrens, 2015). A model with a high score has learned to classify the data better than one with a lower score.

Results

A total of 35 healthy volunteers participated in the two studies. Their ages ranged from 18 to 60 years (average 28.54 years). This database includes the measurements of 19 men and 16 women with an average BMI of 22.9.

As stated before, the sensors were placed on different body parts and the recognizable activities depend on the sensor location. With the sensor placed on the wrist or upper arm, the recognizable activities are cycling, jogging, walking and as one group, lying/sitting/standing. If the sensor is worn on hip, chest or ankle, then the algorithm can detect lying, cycling, jogging, walking and as one group, sitting/standing. If the sensor is placed on thigh, the activities that can be detected are standing, cycling, jogging, walking and as a group sitting/lying.

Table 1 shows the performance of the 2D-CNN. The sensor positioned on the thigh provided the best accuracy, with a value of 99.28% and CK 0.99. In general, it is difficult for the algorithms to distinguish cycling from sitting in some windows. Although during cycling-labeled windows the participants were on the bicycle the whole time, there were periods when they were not peddling. The algorithms find it hard to classify these durations as cycling and not sitting. The worst accuracy occurred with the sensor positioned on the ankle, resulting in an accuracy of 96.57% and CK 0.96. The majority of problems occur during the lying sequences. Both sitting and lying involve minimal acceleration signals, making it difficult to distinguish

Table 1 Activity recognition performance for each sensor location

Sensor Location	Activity	Sensitivity/ Specificity	Accuracy	Cohen's κ
Ankle	Walking	0.98/0.99	0.9657	0.96
	Jogging	0.98/0.99		
	Cycling	0.98/0.999		
	Sitting/Standing	0.93/0.999		
	Lying	0.999/0.98		
Thigh	Walking	0.996/0.995	0.9928	0.99
	Jogging	0.998/0.999		
	Cycling	0.98/0.998		
	Sitting/Lying	1/0.999		
	Standing	0.95/0.999		
Hip	Walking	0.99/0.99	0.9741	0.97
	Jogging	0.98/0.998		
	Cycling	0.98/0.998		
	Sitting/Standing	0.96/0.99		
	Lying	0.96/0.99		
Chest	Walking	0.995/0.99	0.9787	0.97
	Jogging	0.99/0.998		
	Cycling	0.98/0.998		
	Sitting/Standing	0.96/0.99		
	Lying	0.96/0.99		
Wrist	Walking	0.98/0.99	0.984	0.98
	Jogging	0.995/0.998		
	Cycling	0.98/0.99		
	Lying/Sitting/Standing	0.99/0.995		
Upper arm	Walking	0.98/0.99	0.9814	0.97
	Jogging	0.995/0.997		
	Cycling	0.98/0.99		
	Lying/Sitting/Standing	0.98/0.99		

between them so the network categorizes theses sequences using angular velocity. The misclassifications happen when the participant stretches the legs forward or puts them on a table or chair, creating similar conditions and angles to lying on the back. Further information about the results can be found in the supplemental material.

Discussion

In this article we present a 2D-CNN that recognizes human activity by using a wearable sensor. Studies (Cho & Yoon, 2018; Ronao & Cho, 2016; Jiang & Yin, 2015) have shown that accuracies up to 97% can be achieved when using sophisticated deep learning methods like CNN to recognize human activity. In our case, depending on sensor location the HAR accuracy varies from 96.57 to 99.28%, increasing the previous accuracy benchmark by approximately 2%. We also introduce cycling as an additional activity and find that this HAR approach is able to detect cycling with both high sensitivity 0.98 (thigh) and specificity 0.998 (thigh).

The HAR accuracy when the sensor is not placed on the thigh leaves room for improvement and there is still a lot of optimization potential regarding network architecture for HAR algorithms. A way to increase the prediction accuracy is to search for the best hyperparameters, which are the parameters used for training the network, such as the rate at which it learns. In order to achieve better results, it may also be helpful to create a framework that automatically tests different model architectures and varies their hyperparameters.

In this study we used a database with fully or semistandardized movements. Therefore, the network can misclassify some natural movements like sitting with legs stretched forward when the sensor is worn on the ankle. The similarity of the studies presents a chance that the network overfits to the specific settings of the studies. In future we aim to expand our database to include real-life data obtained in every day settings, broadening the use-case of the algorithm. All participants were healthy and their age varied from 18 to 60 years. Future volunteer groups could include a more diverse range of ages.

A larger and richer database could also help the accuracy of CNN since recognizing human activity is a complex problem that starts with a large, diverse, well-structured dataset. The CNN model learns from the data it is provided, and its performance depends on the quality of this database.

Conclusion

Human activity recognition (HAR) is a challenging problem that has recently attracted greater interest in the research community. In this work, we present a two-dimensional convolutional neural network (2D-CNN) capable of classifying activities without the need for feature extraction. We built models for six different sensor locations and distinguish six activities, singled out or grouped, by only using the triaxial accelerometer and the triaxial gyroscope within the Move 4 sensor.

The test data results show that our proposal performs very well and achieves accuracy as high as 99.28% and a Cohen's κ of 0.99 when the Move 4 sensor is worn on the thigh.

Future work includes using more data for network training, considering more complex human activities, finding associations and relationships to health issues for common physical diseases (Avilés-Cruz et al., 2019) to enable real-time HAR directly on the sensor device.

Corresponding address

Eni Hysenllari
movisens GmbH
Karlsruhe, Germany
eni.hysenllari@movisens.com

Acknowledgements. We thank Marco Giurgiu, KIT Karlsruhe and his team for providing us with the data from the study (Giurgiu et al., 2020).

Declarations

Conflict of interest. E. Hysenllari, J. Ottenbacher and D. McLennan disclosed that they are employed by movisens GmbH, which sells the accelerometer devices that are mentioned in the article and analysis software.

For this article no studies with human participants or animals were performed by any of the authors. All studies mentioned were in accordance with the ethical standards indicated in each case.

References

Avilés-Cruz, C., Ferreyra-Ramírez, A., Zúñiga-López, A., & Villegas-Cortéz, J. (2019). Coarse-fine convolutional deep-learning strategy for human activity recognition. *Sensors*, *19*(7), 1556.

Buchner, D. M. (2014). The development and content of the 2008 physical activity guidelines for Americans. *Journal of physical education, recreation and dance*, *85*(7), 13–16.

Cho, H., & Yoon, S. M. (2018). Divide and conquer-based 1D CNN human activity recognition using test data sharpening. *Sensors*, *18*(4), 1055.

Gadri, S., & Neuhold, E. (2020). Building best predictive models using ML and DL approaches to categorize fashion clothes. In *International conference on artificial intelligence and soft computing* (pp. 90–102). Cham: Springer.

Giurgiu, M., Bussmann, J. B., Hill, H., Anedda, B., Kronenwett, M., Koch, E. D., & Reichert, M. (2020). Validating accelerometers for the assessment of body position and sedentary behavior. *Journal for the Measurement of Physical Behaviour*, *3*(3), 253–263.

Hamm, J., Stone, B., Belkin, M., & Dennis, S. (2012). Automatic annotation of daily activity from smartphone-based multisensory streams. In *International conference on mobile computing, applications, and services* (pp. 328–342). Berlin, Heidelberg: Springer.

Ji, X., Cheng, J., Feng, W., & Tao, D. (2018). Skeleton embedded motion body partition for human action recognition using depth sequences. *Signal Processing*, *143*, 56–68.

Jiang, W., & Yin, Z. (2015). Human activity recognition using wearable sensors by deep convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 1307–1310).

O'Shea, K., & Nash, R. (2015). *An introduction to convolutional neural networks*. arXiv preprint arXiv:1511.08458.

Qi, X., Keally, M., Zhou, G., Li, Y., & Ren, Z. (2013). AdaSense: adapting sampling rates for activity recognition in body sensor networks. In *2013 IEEE 19th Real-Time and Embedded Technology and Applications Symposium (RTAS)* (pp. 163–172). IEEE.

Rana, J. B., Shetty, R., & Jha, T. (2015). *Application of machine learning techniques in human activity recognition*. arXiv preprint arXiv:1510.05577.

Ronao, C. A., & Cho, S. B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert systems with applications*, *59*, 235–244.

Shakya, S. R., Zhang, C., & Zhou, Z. (2018). Comparative study of machine learning and deep learning architecture for human activity recognition using accelerometer data. *Int. J. Mach. Learn. Comput*, *8*(6), 577–582.

Staudenmayer, J., Zhu, W., & Catellier, D. J. (2012). Statistical considerations in the analysis of accelerometry-based activity monitor data. *Med Sci Sports Exerc.*, *44*(Suppl 1), 61–67. https://doi.org/10.1249/MSS.0b013e3182399e0f.

Warrens, M. J. (2015). Five ways to look at Cohen's kappa. *Journal of Psychology & Psychotherapy*, *5*(4), 1.

Zhang, C., Yang, X., Lin, W., & Zhu, J. (2012). Recognizing human group behaviors with multi-group causalities. In *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology* (Vol. 3, pp. 44–48). IEEE: In.

Zhou, Z., Li, K., & He, X. (2015). Recognizing human activity in still images by integrating group-based contextual cues. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 1135–1138).