2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Bathroom activities monitoring for older adults by a wrist-mounted accelerometer using a hybrid deep learning model*

Meng Shang[1,2], Yiyuan Zhang[1,2], Ahmed Youssef Ali Amer[1,2,4]
Ine D'Haeseleer[1,3], Bart Vanrumste[1,2]

***Abstract*— Monitoring activities of daily life (ADLs) allows to evaluate health conditions for older adults. However, there are still a limited number of studies on bathroom activities monitoring using a wrist-mounted accelerometer. To fill this gap, in this study, researchers collected data from 15 older adults wearing a wrist-mounted accelerometer. Six bathroom activities, i.e., *dressing*, *undressing*, *brushing teeth*, *using toilet*, *washing face*, and *washing hands*, were investigated. In total, 49.4-hour data for bathroom activities were collected. A hybrid convolutional neural network (CNN) is introduced for bathroom activity recognition. This hybrid CNN model is developed using both hand-crafted and CNN-based features as input. The proposed hybrid CNN model is compared to four machine learning models, i.e., Multilayer Perceptron (MLP), Support Vector Machines (SVM), K-nearest Neighbors (KNN), and Decision Trees (DT), and a conventional CNN model. Based on the classification results of leave-one-subject-out cross-validation (LOSO), the hybrid CNN model outperformed the other models. The hybrid CNN model is also tested based on a transfer learning method. As a calibration step based on LOSO, the transfer learning method additionally trains the model with an example of each activity from the test subject. The transfer learning method obtained better classification performance than LOSO. With transfer learning, the f1-score for *using toilet* was improved from 0.7784 to 0.8437. This study proposes a deep learning model fusing hand-crafted features and CNN-based features. Besides, the transfer learning method offers a way to build subject-dependent models to improve the classification performance.**

***Clinical relevance* — This provides a model that helps monitoring older adults' bathroom activities using a single wrist-mounted accelerometer.**

I. INTRODUCTION

Many older adults suffer from some age-related diseases such as cardiovascular diseases and cancers [1]. To prevent or treat such diseases, caregivers need to monitor and control older adults' activities of daily life (ADLs). Common ADLs include personal hygiene, dressing, toileting, ambulating, etc. [2]. A traditional method of ADLs monitoring is using self-reports [3]. However, older adults could not recall ADLs accurately after some weeks or months.

With the development of microelectronics, the drawbacks of self-reports could be overcome by using sensors. The sensors used in previous studies were often wearable devices (e.g., inertial sensors) or ambient devices (e.g., passive infrared sensors) [4]. Compared with ambient sensors, the advantages of wearable sensors are low cost and ease of manipulation. Therefore, in this study, a wearable accelerometer is used for collecting the movement signals.

Some bathroom activities have been studied via a single accelerometer [5] or multiple sensors [6], [7]. However, some activities still cannot be recognized accurately using a single wrist-mounted accelerometer. The data collected in this study has been processed by another study [5], which proposed a shallow neural network and proved it possible to recognize *using the toilet* and non-bathroom activities.

For traditional machine learning methods, hand-crafted features are extracted to represent characteristics of the raw data [6]. On the other hand, deep learning models such as Convolutional Neural Network (CNN) could extract features by themselves [8], [9]. Although CNN is more applied in image classification, it has already been proved to be useful in processing sensor signals [8]. The features extracted by CNN can also be fused with the hand-crafted features to improve classification performance [9]. Additionally, transfer learning could be applied for training machine learning models to further reduce inter-subject confusion [10].

This study aims to find the models leading to the best classification results for monitoring bathroom activities. This paper proposes a hybrid CNN model to recognize bathroom activities using hand-crafted features and raw data from the accelerometer. This hybrid CNN model is compared with traditional machine learning models which only applied hand-crafted features and a conventional CNN model. Transfer learning technology is also applied to build subject-dependent models via including an example of each activity of the test subject for training.

II. METHODOLOGY AND MATERIALS

A. Data acquisition

An Empatica E4 wristband [11] was used for collecting data. This is a wristband integrated with a 3-axis accelerometer, with a sampling frequency of 32 Hz and the detectable range of $\pm 2g$. In total, 15 older adults (aged 76 ± 10 years old) were recruited. Seven of them were males and eight of them were females.

Subjects were asked to wear the wristband on the wrist of the dominant-hand and annotate six bathroom activities:

*This research study was possible with the support of KIC EIT Health funding for GRaCE- AGE.

[1]The authors are with KU Leuven, e-Media Research Lab, 3000 Leuven, Belgium. {meng.shang, yiyuan.zhang, ahmed.youssefaliamer, ine.dhaeseleer, bart.vanrumste}@kuleuven.be

[2]The authors are with KU Leuven, Department of Electrical Engineering

[3]The authors are with KU Leuven, Department of Computer Science

[4]The author is with KU Leuven, Measure, Model & Manage Bioresponses (M3-BIORES), Department of Biosystems

dressing (dress), *undressing* (undress), *brushing teeth* (teeth), *using toilet* (toilet), *washing face* (face), and *washing hands* (hand). Some other ADLs were also annotated as *others*. A tablet application was developed and used by the subjects to annotate the bathroom activities. The data was collected for two consecutive weeks. Six bathroom activities accounted for 49.4 hours in total. For the *others* activity, only 75-hour data was kept in total while the rest was discarded. The reason was that there were too many wrong annotations since the subjects less followed the instructions to annotate activities with the past of time [5].

Ethical approval was obtained by UZ Leuven Ethical Committee, trial number S60250.

B. Signal pre-processing

Signal pre-processing included three stages: filtering and magnitude extraction.

1) Filtering: Two Butterworth low-pass filters were applied to the original acceleration signals. One filter with a cutoff frequency of 6 Hz was used to remove high-frequency noise to obtain the de-noised component (ax, ay, and az). Then the other filter with a cutoff frequency of 0.45 Hz was used to separate the de-noised component to the gravity component (ax_gravity, ay_gravity, az_gravity, in the frequency range of 0 Hz- 0.45 Hz) and the movement component (ax_movement, ay_movement, az_movement, in the frequency range of 0.45 Hz- 6 Hz) [5].

2) Magnitude extraction: The magnitude of the de-noised component (M), gravity component (M_gravity), and movement component (M_movement) was calculated separately according to the equation:

$$M = \sqrt{ax^2 + ay^2 + az^2} \tag{1}$$

C. Hand-crafted features extraction

Features were extracted from each component, using a sliding window of 32s (1024 samples) with 50% overlap. In total, there were 29 types of features extracted from the time-domain and frequency-domain [4]. In the time-frequency-domain, the detailed coefficients of the third to sixth levels of the wavelet transform were included [12], using Daubechies wavelet (db5). The Euclidean norm and the squared Euclidean norm of the coefficients of each level were calculated as features. The age and gender of each subject were also included as features. In total, 407 features were extracted.

D. Classification models

Four traditional machine learning models and two deep learning models were applied. The hyper-parameters of the traditional models were optimized by Random Search algorithm. The traditional models [4] and corresponding hyper-parameters (searching results, searching space) were:

- Support Vector Machine (SVM)
- Multilayer perception (MLP) (hidden layers=3, [1:1:10])
- K-nearest neighbor (KNN) (K=3, [3:2:35])
- Decision tree (DT) (maximum depth=33, [1:1:35])

Deep learning models included CNN and Hybrid CNN. The hybrid CNN model was applied to concatenate hand-crafted features and CNN-based features. The proposed hybrid CNN model is shown in Fig. 1. The de-noised components (ax, ay, az) were applied as input. As applied in hand-crafted features extraction, these components were also segmented using a sliding window of 32s with 50% overlap. Therefore, the input size was 1024*3. The CNN-based features were extracted by three convolutional layers. Each convolutional layer was followed by a max-pooling layer. Then, these features were flattened and concatenated with hand-crafted features. These fused features were fed into three fully connected layers. To compare with the hybrid CNN model, another CNN model without hand-crafted features was also applied. The structure of the convolutional layers was the same as the ones in [**?**], since it has been applied to a larger dataset from a single accelerometer. In this study, these layers were trained using our dataset. The deep learning models were developed using GPU, with the training batch size of 32.

E. Training and testing methods

The models were trained in two scenarios: 1) leave-one-subject-out cross-validation (LOSO) and 2) LOSO and transfer learning.

1) LOSO: Firstly, the six models were all trained and evaluated by LOSO. In this scenario, the data of one subject was applied as the testing dataset, while the data of the other 14 subjects was applied for training the model. This was aimed to evaluate the classification performance of the model in new subjects.

2) LOSO and transfer learning: Then, the transfer learning method was applied to the best-performed model in the first scenario. In this method, the model trained in LOSO was used as the pre-trained model. Afterward, an example of each activity event from the new subject (target subject) was used to continue training the pre-trained model for calibration. Finally, the rest of the examples were used to test the model. These models were subject-dependent since the pre-trained model needs to be trained by a part of data from each single target subject.

The performance of models was evaluated by the f1-score.

III. RESULTS

A. Classification performance of LOSO

Table I shows the f1-scores of six models based on LOSO, along with the numbers of parameters (Np) of some models. The hybrid CNN obtains the highest f1-scores for all bathroom activities. *Using toilet* and *brushing teeth* have the highest f1-scores of 0.7784 and 0.6620, respectively.

The confusion matrix of the Hybrid CNN model is shown in Fig. 2. The main confusion occurs between *others* and each of six bathroom activities. On the other hand, within the bathroom activities, the confusion occurs among *dressing*, *undressing* and *washing face*. To show the performance variance among subjects, Fig. 3 gives the box plots of f1-scores for the 15 subjects using the hybrid CNN model. Big

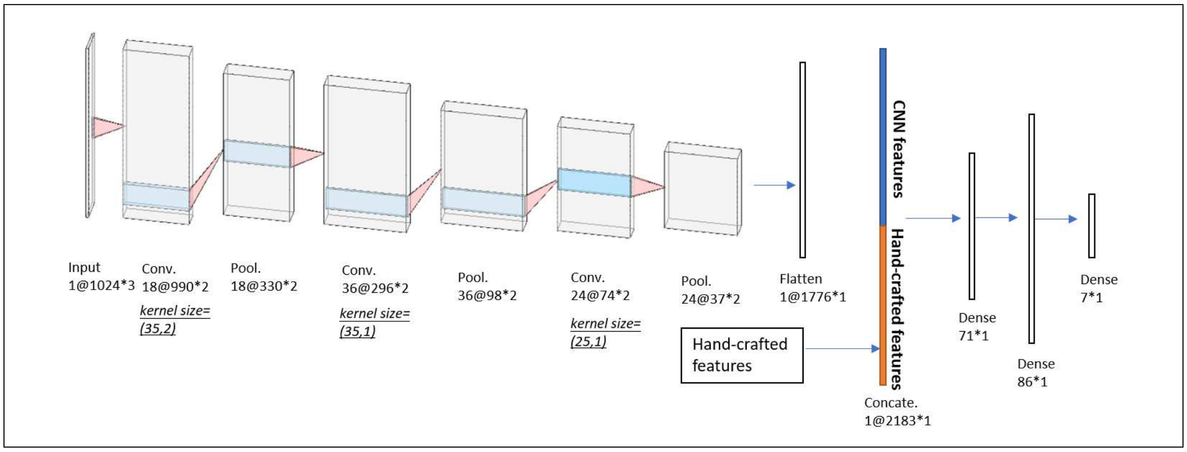

Fig. 1. The architecture of proposed hybrid CNN model. Each pooling layer has a pool size of (3,1) and a stride of 3

TABLE I

F1-SCORES AND NP OF SIX MODELS EVALUATED BY LOSO WITH AND WITHOUT TRANSFER LEARNING

	dress	face	toilet	hand	undress	teeth	Np
LOSO without transfer learning							
MLP	0.3527	0.2595	0.7654	0.2301	0.2874	0.6208	19941
SVM	0.2452	0.1903	0.6988	0.2027	0.2199	0.5212	\
KNN	0.2001	0.1729	0.6552	0.1623	0.1788	0.4835	\
DT	0.1933	0.1684	0.6421	0.1573	0.1687	0.4758	\
CNN	0.1567	0.1474	0.5899	0.1243	0.1395	0.4129	136176
hybrid CNN	0.4525	0.4083	0.7784	0.2485	0.4003	0.6620	165783
LOSO with transfer learning							
hybrid CNN	**0.6775**	**0.6600**	**0.8437**	**0.7368**	**0.7160**	**0.8243**	165783

variance occurs in each bathroom activity and there is a zero f1-score in *undressing* and *washing hands*.

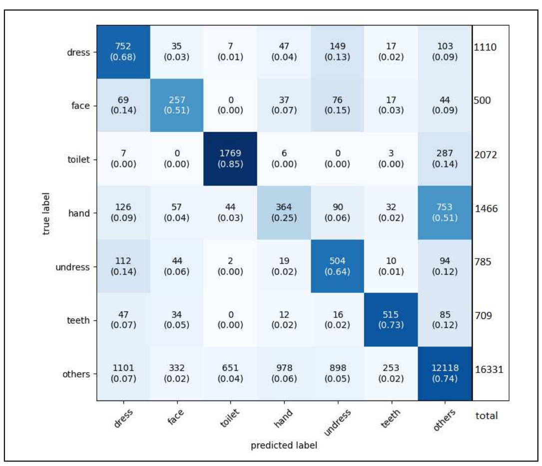

Fig. 2. The confusion matrix of the hybrid CNN mode based without transfer learning. The normalized confusion matrix over true rows are shown in the brackets. Total numbers of true labels are shown in the right column.

B. *Classification performance of transfer learning*

To compare the results of LOSO and transfer learning, Table I gives the f1-scores of the Hybrid CNN model with and without transfer learning. The classification performance is improved with transfer learning as shown in bold. Compared with LOSO, the most improved class is *washing hands* (with f1-scores from 0.2485 to 0.7368). The f1-scores of *using toilet* and *brushing teeth* increase to 0.8437 and 0.8243, respectively. The confusion matrix of transfer learning using the hybrid CNN model is shown in Fig. 4. The confusion between bathroom activities and *others* is improved compared with LOSO in Fig. 2.

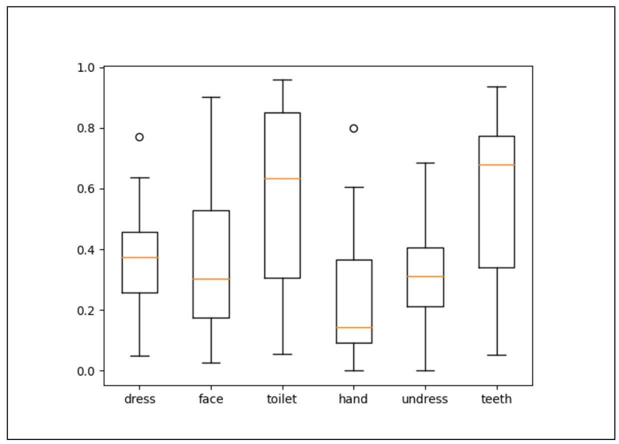

Fig. 3. Box plots of f1-scores for the 15 subjects using the hybrid CNN model without transfer learning

IV. DISCUSSION

A. *Models comparison*

The MLP, SVM, KNN, and DT have hand-crafted features as input while the hybrid CNN model fuse the hand-crafted features and CNN-based features. The results show that the CNN model obtains the best f1-scores for all activities. The combination of the hand-crafted features and CNN-based features could give a more complicated description of data that is hard for humans to interpret. Especially for subject-independent models, CNN-based features indicate more general information that could be applied to unseen data, as concluded by [13].

Although involving CNN-based features improves the classification results, the CNN model without hand-crafted

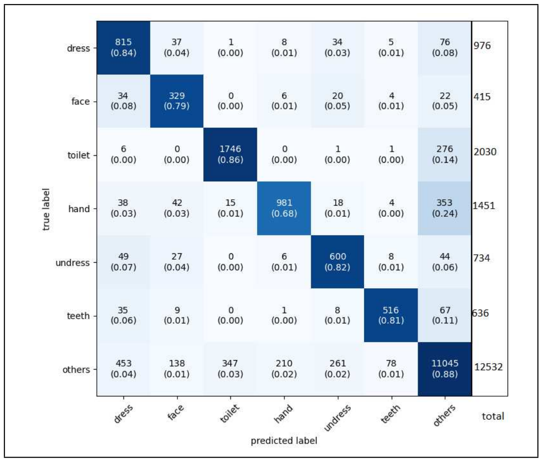

Fig. 4. The confusion matrix of the hybrid CNN mode based on transfer learning. The normalized confusion matrix over true rows are shown in the brackets. Total numbers of true labels are shown in the right column.

features has the lowest f1-scores. Future studies could focus on developing more types of deep learning models.

B. Activity recognition

This part will only discuss the activity classification results of the hybrid CNN model. Since LOSO and transfer learning have similar confusion among activities, they will be discussed together.

Among bathroom activities, *dressing*, *undressing*, and *washing face* confuse with each other. This confusion could be caused by the fact that *dressing* and *undressing* include arm lifting behavior and *washing face* also requires subjects to lift arm to use a towel.

For all bathroom activities, the main confusion comes from *others*. This activity may include similar arm movements to bathroom activities. For example, in the kitchen, there could be activities like *washing dishes* and *washing vegetables* which can be confused with *washing hands*. However, from the other activities, we do not have detailed annotations. Besides, the small sample size makes the generalization results less reliable. In the future, additional environmental sensors could be applied to offer more information. It has been concluded that fusing features from different sensors could improve classification performance [6].

C. Transfer learning

In LOSO, as shown in Fig. 3, the results show that the classification performance among subjects has a large variance. Since the dataset was collected in real environments annotated by subjects themselves, this performance variance may be caused by their different behaviors and habits, i.e., different contexts. This is the disadvantage of subject-independent models.

According to the comparison between the results of LOSO and transfer learning, the subject-dependent models show higher f1-scores. In the future, studies could focus on developing subject-dependent models in real applications.

V. CONCLUSION

This study is aimed to recognize the bathroom activities of older adults using a wearable accelerometer by deep learning methods. Based on the result and discussion, it is recommended to use a hybrid CNN model fusing CNN features and hand-crafted features to reach high classification results. The experiment of transfer learning proves that it is possible to build subject-dependent models to improve classification performance. In future studies, researchers will validate the usage of environmental sensors for distinguishing bathroom activities from other daily activities.

REFERENCES

[1] Z. Milanović, S. Pantelić, N. Trajković, G. Sporiš, R. Kostić, and N. James, "Age-related decrease in physical activity and functional fitness among elderly men and women," Clin Interv Aging, vol. 8, pp. 549–556, 2013.

[2] M. E. Mlinac and M. C. Feng, "Assessment of Activities of Daily Living, Self-Care, and Independence," Archives of Clinical Neuropsychology, vol. 31, no. 6, pp. 506–516, Sep. 2016.

[3] R. S. BUCKS, D. L. ASHWORTH, G. K. WILCOCK, and K. SIEGFRIED, "Assessment of Activities of Daily Living in Dementia: Development of the Bristol Activities of Daily Living Scale," Age and Ageing, vol. 25, no. 2, pp. 113–120, Mar. 1996, doi: 10.1093/ageing/25.2.113.

[4] Y. Wang, S. Cang, and H. Yu, "A survey on wearable sensor modality centred human activity recognition in health care," Expert Systems with Applications, vol. 137, pp. 167–190, 2019.

[5] Y. Zhang, J. Wullems, I. D'Haeseleer, V. Vanden Abeele and B. Vanrumste, "Bathroom activity monitoring for older adults via wearable device," 2020 IEEE International Conference on Healthcare Informatics (ICHI), Oldenburg, Germany, 2020, pp. 1-10.

[6] S. Chernbumroong, S. Cang, A. Atkins, and H. Yu, "Elderly activities recognition and classification for applications in assisted living," Expert Systems with Applications, vol. 40, no. 5, pp. 1662–1674, Apr. 2013.

[7] N. F. Ince, C.-H. Min, A. Tewfik, and D. Vanderpool, "Detection of Early Morning Daily Activities with Static Home and Wearable Wireless Sensors," EURASIP J. Adv. Signal Process., vol. 2008, no. 1, p. 273130, Dec. 2007.

[8] Y. Chen and Y. Xue, "A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer," in 2015 IEEE International Conference on Systems, Man, and Cybernetics, Oct. 2015, pp. 1488–1492, doi: 10.1109/SMC.2015.263.

[9] M. Dong, J. Han, Y. He, and X. Jing, "HAR-Net: Fusing Deep Representation and Hand-Crafted Features for Human Activity Recognition," in Signal and Information Processing, Networking and Computers, Singapore, 2019, pp. 32–40.

[10] R. Saeedi, K. Sasani, S. Norgaard, and A. H. Gebremedhin, "Personalized Human Activity Recognition using Wearables: A Manifold Learning-based Knowledge Transfer," in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jul. 2018, pp. 1193–1196.

[11] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti, "Empatica E3 — A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition," in 2014 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH), Nov. 2014, pp. 39–42.

[12] J. Mantyjarvi, J. Himberg, and T. Seppanen, "Recognizing human motion with multiple acceleration sensors," in 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat. No. 01CH37236), 2001, vol. 2, pp. 747–752.

[13] G. Antipov, S.-A. Berrani, N. Ruchaud, and J.-L. Dugelay, "Learned vs. Hand-Crafted Features for Pedestrian Gender Recognition," in Proceedings of the 23rd ACM international conference on Multimedia, New York, NY, USA, Oct. 2015, pp. 1263–1266.