Evaluating Movement and Device-Specific DeepConvLSTM Performance in Wearable-Based Human Activity Recognition

Gabriela Ciortuz[a], Hawzhin Hozhabr Pour[b] and Sebastian Fudickar[c]

Institute of Medical Informatics, University of Lübeck, Ratzeburger Allee 160, Lübeck, Germany

Keywords: Human Activity Recognition, Wearables, Deep Learning, Neural Networks, Time-Series, Behaviour, Functional Assessment, Physical Condition.

Abstract: This article provides a comprehensive look at human activity recognition via three consumer devices with different body placements and a deep hybrid model containing CNN and LSTM layers. The used dataset consists of 53 activities recorded from the motion sensors (IMUs) of the three devices. Compared to the available human activity recognition datasets, this dataset holds the biggest number of classes, enabling us to provide an in-depth analysis of activity recognition for health-related assessments, as well as a comparison with other benchmark models such as a CNN and LSTM model. In addition, we categorize the activities into six movement groups and discuss their relevance for health-related assessments. Our results show that the hybrid model outperforms the benchmark models for all devices individually and all together. Furthermore, we show that the smartwatch could as a standalone consumer device classify activities in the six movement groups very well and for most of the use cases using a smartwatch would be practical.

1 INTRODUCTION

The interest in wearable-based Human Activity Recognition (HAR) has rapidly grown because of its ability to monitor health and well-being indicators of individuals as digital biomarkers (Mekruksavanich et al., 2020). HAR research has mainly focused on wearable-based approaches, as they have been preferred over camera systems due to privacy concerns (Uddin and Soylu, 2021). Moreover, consumer devices, such as smartphones and smartwatches are easily accessible, omnipresent and have a high user acceptance (Dave et al., 2022). Considering the omnipresence of smartphones, it is a logical step to leverage data obtained from smartphone sensors for continually collecting data in different contexts (Friedrich et al., 2019). Other consumer devices, such as smartwatches and smart glasses can also enable the continuous monitoring of daily activities beyond the confines of personal living spaces, unlocking opportunities for a more thorough comprehension of people's health, contributing especially to the early detection of physical condition degradation, timely diagnosis and prognosis of health issues (Dave et al., 2022).

In addition, based on the foundation of HAR, corresponding health quality indicators, such as power measurements and functional assessment parameters can be extracted (Hellmers et al., 2018). Depending on the health-related problem, different activities are relevant. In the context of chronic disease monitoring, for example in heart disease or diabetes, accelerometers can provide information about the total amount, intensity and duration of daily physical activity (Hans Van Remoortel et al., 2012). HAR helps in monitoring patients' exercise routines, degree of mobility, and compliance with prescribed activities.

Many methods have been proposed for HAR, starting with traditional machine learning algorithms, all the way to Deep Learning (DL) models, such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Transformer-based and deep hybrid models, containing multiple types of layers (Li et al., 2018; Augustinov et al., 2023).

Comparison studies have highlighted the effectiveness of hybrid deep learning models over feature learning methods for HAR (Li et al., 2018). One popular hybrid model which has become state-of-the-art for HAR is the DeepConvLSTM (Ordóñez and Roggen, 2016). This DeepConvLSTM hybrid architecture has produced outstanding results in many

[a] https://orcid.org/0000-0001-9443-7825
[b] https://orcid.org/0000-0003-4404-7313
[c] https://orcid.org/0000-0002-3553-5131

Ciortuz, G., Hozhabr Pour, H. and Fudickar, S.
Evaluating Movement and Device-Specific DeepConvLSTM Performance in Wearable-Based Human Activity Recognition.
DOI: 10.5220/0012471300003657

In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 2*, pages 746-753
ISBN: 978-989-758-688-0; ISSN: 2184-4305

open-access datasets, outperforming other deep learning models (Bock et al., 2022; Aboo and Ibrahim, 2022).

Most of the existing HAR studies usually deal with the classification of five and up to 27 classes (Zhang et al., 2022). Available literature either focused on a specific use case with a limited number of activities of interest or used all the labelled activities in publicly available datasets to prove that their models worked for the classification task (Gil-Martín et al., 2020).

As a result, it remains uncertain, if the DeepConvLSTM can classify very various types of activities and scale up to classify between a large number of classes. To close this gap, in our article, we train the CogAge dataset collected in (Li et al., 2020) using the DeepConvLSTM. The CogAge dataset has 53 classes, making it the biggest HAR dataset known in terms of the number of activity labels. This fact makes it the ideal dataset for performing an in-depth evaluation of the classification results for three consumer devices - a smartphone, a smartwatch and smart glasses.

The main contributions of this article are:

- We train and evaluate a DeepConvLSTM model using a dataset of 53 activities.
- We analyse and compare the classification results of the DeepConvLSTM as benchmark models such as CNN and LSTM.
- We experiment with data from three consumer devices placed on three different body parts, highlighting the health-related use cases for each sensor.
- We propose a new classification of activities into six movement groups of interest for health-related assessments.

2 BACKGROUND AND RELATED WORK

The following two main aspects in HAR research are usually of great importance: Which type of activities are to be classified and what algorithms perform that classification task the best?

2.1 Types of Activities in HAR

Many categorisations of human activities have been suggested, that facilitate the development and evaluation of HAR models especially for video-based HAR (Hussain et al., 2020). For wearable-based HAR, the main categories of activities are atomic activities, for which (Gil-Martín et al., 2020) proposed classifying them into postures, gestures and repetitive actions and complex activities such as Activities of Daily Living (ADLs) (Nisar et al., 2020). Atomic activities are short-term, simple actions (Morshed et al., 2023), which combined in longer sequences make up an ADL, which is generally a complex task (Nisar et al., 2020).

Our as well as other studies/datasets (Li et al., 2020; Roggen et al., 2012) categorized atomic activities based on state or locomotion and behavioural activity types. State or locomotion activities are related to the posture or means of moving from one place to another (Roggen et al., 2012). Behavioural activities characterize the actions or the behaviour of the subject (Li et al., 2020). The latter activities are more difficult to recognize because of multiple factors: Firstly, many more behavioural activities can be defined than state activities, and secondly, they can be very similar in their movement pattern.

In our paper, we focus on behavioural atomic activities. We leave the state activities out of our analysis because they have already been analysed in-depth and consumer devices can easily recognize such activities by employing simple models, such as CNNs. Recognizing more complex activities either demands a substantial amount of training data or some Recurrent Neural Networks (RNN) that process long-term dependencies to enable the learning of intricate features (Abbaspour et al., 2020).

2.2 DeepConvLSTM

Hybrid CNN-LSTM models have been used very often in HAR because they combine the powers of both CNN and LSTM layers to process sequential data (Ordóñez and Roggen, 2016). By employing CNN layers, the model can extract spatial features. Additionally, the LSTM layers can capture temporal changes from raw sensor data (Roy et al., 2023). (Ordóñez and Roggen, 2016) first proposed the DeepConvLSTM. (Bock et al., 2022) proposed that a more shallow architecture.

The CNN-LSTM architecture is comprised of two 1D CNN layers at first, followed by a dropout layer, a 1D MaxPooling and a flattening layer, which are all wrapped in a time-distributed layer, allowing for the same layers to process all the subsequences in the window. After the processed data is flattened, it is fed to a stack of LSTM layers before a dropout and two dense layers, of which the last layer uses a softmax activation for the classification (Phyo and Byun, 2021). Figure 1 depicts the model architecture.

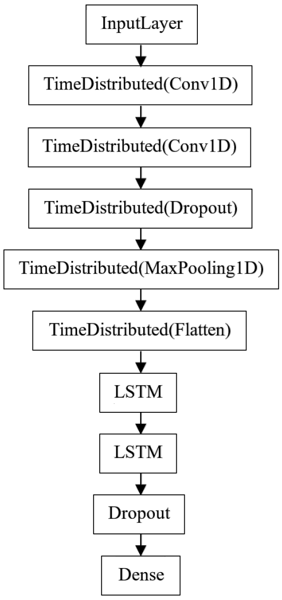

Figure 1: The architecture of the DeepConvLSTM model used in our experiments.

3 METHODS

In the following subsections, the methods and the dataset used in our study are introduced: The dataset used, the preprocessing steps applied before the training of the neural network as well as the experiments that we conducted.

3.1 Dataset Description

We used a publicly available dataset of behavioural atomic activities collected from four subjects in (Li et al., 2020) with three devices: A Google NEXUS 5X smartphone, a Microsoft Band 2 and Jins MEME glasses placed on the body of the subjects.

The subjects were asked to perform the 55 activities at least 20 times in two sessions recorded on two different days. The three devices measured data using different sampling rates and also had different sensor channels. The smart glasses provided data from two sensor modalities: Three-axis accelerometer and gyroscope sampled at 20 Hz. The smartwatch provided the same two sensor modalities as the smart glasses but sampled at 67 Hz. The smartphone data comes from four sensor modalities all sampled at 200 Hz: Three-axis accelerometer, gravity, gyroscope, and linear accelerometer.

The original CogAge dataset from (Li et al., 2020) has 55 classes for two configurations: One configuration where both hands were used for the activities and one configuration, where only the left hand, where the smartwatch was on, was used. Since we wanted to differentiate between the various activities, we chose the left-hand-only configuration and removed two labels, for which there were no measurements done with the left hand only, leading to 53 labels used in our study. In contrast to other prominent public datasets for HAR using wearable data, the CogAge dataset has by far the highest number of classes. The three datasets with the highest number of classes reported in (Zhang et al., 2022) have 27, 19 and 18 classes, respectively. Consequently, we have at least twice as many classes as these commonly used public datasets. This underscores the suitability of the CogAge dataset for our study's specific scope, allowing us to comprehensively assess the model's performance across the three devices and a wide array of behavioural activities.

For the training and testing of the models, we used a subject-dependent split of the dataset, where the data from a first recording session were used for training, and the data from the second recording session for testing, ensuring an almost equal split between training and testing data. Figure 2 depicts the 53 activities we used and the number of segments of each activity in the training and testing dataset.

3.2 Preprocessing

Before feeding the data into the deep neural network we first applied a Z-score normalization Secondly, similarly to other works (Irshad et al., 2022; Mahmud et al., 2020) using multiple devices with different sampling rates, we resampled the data to achieve a better comparison of the results of the models trained on the devices individually and on all the devices together. We upsampled the data to 200 Hz, which is the highest sampling frequency of the three devices. Because upsampling may lead to overfitting, we have also downsampled the data to 20 Hz. This approach did not yield as good results as the upsampling method, because of the loss of too much information, thus we proceeded with the upsampled data.

Furthermore, we modified the data representation to be able to feed it to the hybrid CNN-LSTM model, which uses subsequences as input to the CNN layers. A well-crafted feature representation results in more informative and discriminative features, which in turn contributes to an improvement in overall performance (Ciortuz et al., 2023).

3.3 Deep Models

We established a DeepConvLSTM model designed to process sequences of four seconds duration and a spe-

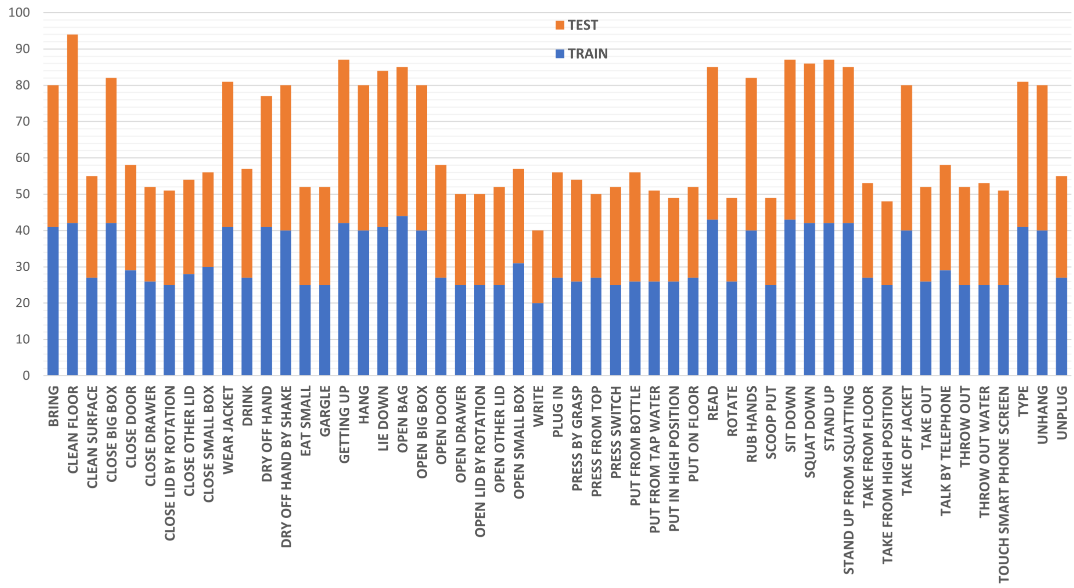

Figure 2: Number of segments of each activity in the dataset used for training and testing.

cific number of features, as required by the respective sensor or combinations of sensors as mentioned in Chapter 3.2.

Using the same sampling frequency for all sensors after upsampling allowed us to build subsequences of the same length to feed to the CNN layers. Considering the factors mentioned above, we segmented the data in subsequences of 160 milliseconds, leading to a total of 25 subsequences per sample. The subsequences are encoded by the CNN layers, flattened and then fed to the LSTM layers.

As benchmark models, we trained a CNN model containing three 1D convolutional layers with MaxPooling, then a flattening layer and a dense layer for the classification. We also trained an LSTM model with one LSTM and one dense layer before a second dense layer for classification. Both models were trained for the same number of epochs as the DeepConvLSTM.

3.4 Experiments

We trained device-specific models for each of the three devices as well as one combined model to analyze the performance in recognizing the different types of activities in the dataset and allow investigation of the benefit of deploying device-specific models.

Each experiment was conducted as follows: For each device and all of them together, we trained and evaluated the models ten times and provided the results in the form of the mean of each metric. We trained the models using the Adam optimiser with a categorical cross-entropy loss function for 1000 epochs. Then we computed the prediction of the test data on the model that achieved the highest validation accuracy during training.

We also employed hybrid ConvLSTM architectures with different depths by adding one to four more LSTM layers to the model. A deeper hybrid architecture increased the complexity of the model without improving the results, similar to the outcomes in (Bock et al., 2022). Consequently, we decided to use a simplified model containing only two LSTM layers.

For the evaluation part, we grouped the activities into six groups of interest for health-related assessments as follows:

1. **Opening/Closing:** Close Big Box, Close Door, Close Drawer, Close Lid By Rotation, Close Other Lid, Close Small Box, Open Bag, Open Big Box, Open Door, Open Drawer, Open Lid By Rotation, Open Other Lid, Open Small Box
2. **Press and Pull:** Plug In, Press By Grasp, Press From Top, Press Switch, Unplug
3. **Raising/Lowering:** Bring, Hang, Put From Bottle, Put From Tap Water, Put In High Position, Put On Floor, Scoop Put, Take From Floor, Take From High Position, Take Out, Throw Out Water, Throw Out, Unhang
4. **Body Movement:** Clean Floor, Getting Up, Lie Down, Sit Down, Squat Down, Stand Up, Stand Up From Squatting, Take Out Jacket, Wear Jacket
5. **Hand to Head:** Drink, Eat Small, Gargle, Talk By Telephone

6. **Continuous Hand/Head Movements:** Clean Surface, Dry Off Hand, Dry Off Hand By Shake, Read, Rotate, Rub Hands, Touch Smart Phone Screen, Type, Write

3.5 Statistical Evaluation

As evaluation metrics of the models, we included the average F1-score (AF1), the accuracy, the Mean Average Precision (MAP) and the Area Under the Curve (AUC).

4 RESULTS

The results in form of the mean of the Average F1 score, Accuracy, MAP and AUC in % of the activity recognition over 10 runs for each of the three devices individually as well as trained together and for each model: CNN, LSTM and DeepConvLSTM are shown in Table 1.

We plotted the class F1 score in each of the groups mentioned above and per each device for a better understanding of their performance. The mean tendency per group can be seen in Table 2.

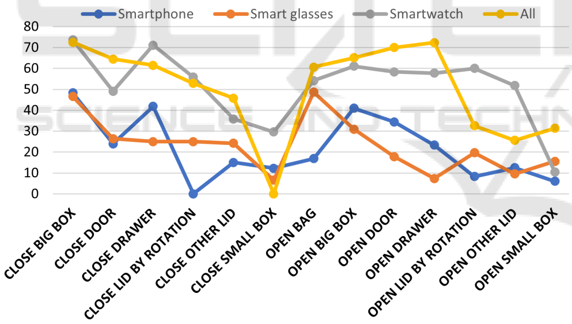

Figure 3: The F1 score per class in group 1 (Closing/Opening).

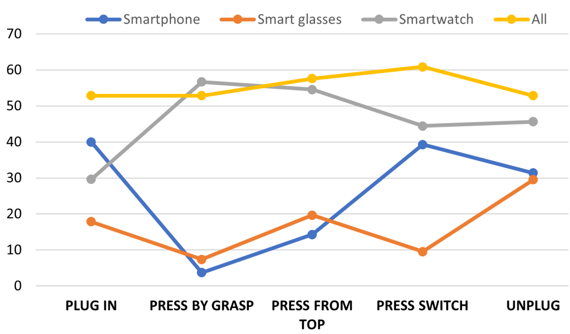

Figure 4: The F1 score per class in group 2 (Press and Pull).

Table 1: The mean of the Average F1 Score, Accuracy, MAP and AUC (in %) of the activity recognition over 10 runs for smartphone, smart glasses and smart watch individually and all together.

	AF1	Acc	MAP	AUC
	Smartphone			
CNN model	23.75	25.55	23.40	82.89
LSTM model	23.34	23.71	20.61	83.01
Deep ConvLSTM	**32.07**	**32.13**	**30.35**	**86.02**
	Smart glasses			
CNN model	28.55	29.09	27.29	86.82
LSTM model	36.85	37.06	34.59	**86.83**
Deep ConvLSTM	**37.07**	**37.37**	**35.67**	86.62
	Smartwatch			
CNN model	48.35	48.54	48.51	91.41
LSTM model	54.08	54.25	54.77	93.41
Deep ConvLSTM	**61.78**	**61.84**	**64.06**	**95.40**
	All devices			
CNN model	62.73	**66.52**	**69.81**	**96.54**
LSTM model	51.45	51.78	52.27	92.78
Deep ConvLSTM	**65.64**	65.65	68.80	96.11

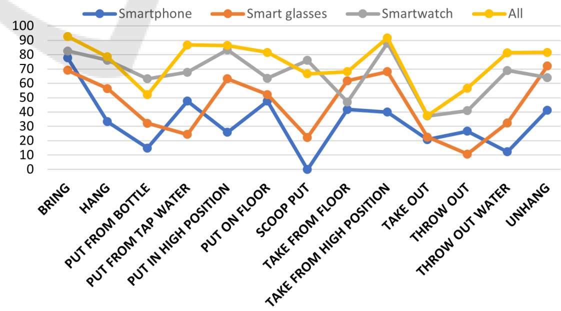

Figure 5: The F1 score per class in group 3 (Raising/Lowering).

5 DISCUSSION

We observe in Table 1 that the DeepConvLSTM models outperform the benchmark CNN and LSTM models, making it an obvious choice of architecture for each sensor or all of them together as well. At the

Table 2: The mean (and standard deviation) of the F1 Score (in %) of the activities per group for the smartphone, smart glasses and smart watch independently as well as all together.

	Group	Nr. of Activities	Smartphone	Smart glasses	Smartwatch	All
1.	Opening/ Closing	13	21.85 (14.66)	23.33 (12.66)	**51.45 (16.55)**	50.37 (21.22)
2.	Press and Pull	5	25.72 (14.39)	16.78 (7.89)	46.17 (9.56)	**55.39 (3.30)**
3.	Raising/ Lowering	13	33.08 (19.05)	45.17 (20.73)	66.04 (15.46)	**73.94 (16.10)**
4.	Body Movement	9	72.94 (19.51)	62.93 (10.53)	74.94 (7.94)	**86.18 (8.93)**
5.	Hand to Head	4	23.74 (16.72)	37.12 (16.72)	**79.23 (4.07)**	76.27 (6.32)
6.	Continuous Hand/ Head Movement	9	16.57 (11.58)	30.01 (11.84)	69.90 (16.29)	**72.37 (13.36)**

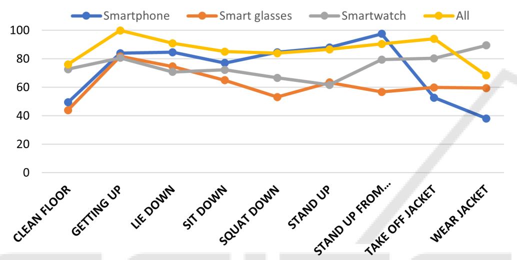

Figure 6: The F1 score per class in group 4 (Body Movement).

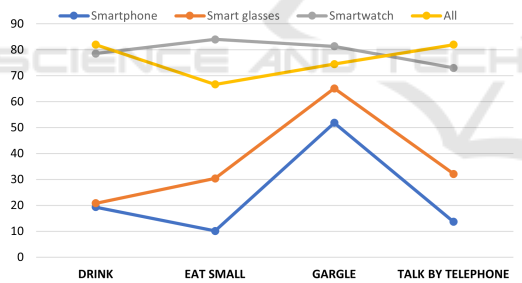

Figure 7: The F1 score per class in group 5 (Hand to Head).

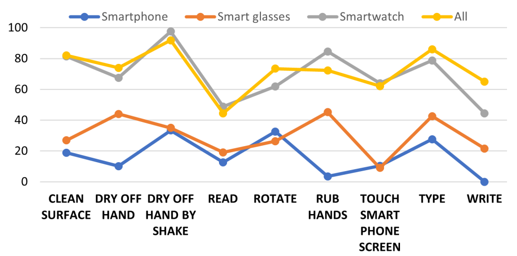

Figure 8: The F1 score per class in group 6 (Continuous Hand/Head Movements).

same time, the same table shows, that the CNN is a better choice over the LSTM, when there are many features and higher sampling frequencies, such as for the smartphone (F1 scores 23.75 % and 23.34 %) and all devices together (F1 scores 62.73 % and 51.45 %). In contrast for the smart glasses and smartwatch, the LSTM model performed much better than the CNN (F1 scores 36.85 % and 28.55 % for the smart glasses and 53.08 % and 48.35 % for the smartwatch). These findings highlight the fact, that the DeepConvLSTM can be generally applied to different HAR classification tasks achieving very good results. The DeepConvLSTM achieved an F1 score between 3% and 14% higher than the other models regardless of the sampling frequency, segment length and number of features in the data.

For the evaluation, we categorized the activities into six groups. The grouping was done by the level of similarity of the activities and then we analysed the relevance of each group for health-related assessments, as outlined below.

In general, the tendency in most groups of similar activities was that considering the data from all sensor devices for training achieved higher classification performance than the device-specific training performance, with two exceptions: Group 1: Closing/Opening where the mean F1 score of the smartwatch individually is 51.45 %, which is over one % higher than the mean F1 score using all the sensor channels (50.37%) and Group 5: Hand to Head, where the smartwatch also outperformed the mean F1 score of all the channels by almost three % (79.23 % and 76.18 %). Since using all of the sensors is not possible in many contexts and because specific devices performed very well, we can argue that one device can be used independently, without many disadvantages. For example, when working with elderly users in a real-life context, using more than one device might be overwhelming for them.

Looking at the first group (Closing/Opening) in Figure 3, containing 13 activities related to opening or closing, we observe that the smartwatch performed very well in classifying them (54.56 % mean F1 score) - see 2. Such activities might be relevant in monitoring rehabilitation after operations or while recovering from cerebrovascular strokes for example.

Group number two contains five activities where a pressing or pulling motion is performed. These are also activities where the hand is involved, thus the smartwatch is suitable and enough for their recognition, achieving an F1 score of 46.17% (Using all the devices, the model classification F1 score was 55.39%) - see Table 2. For the activities of pressing a switch, plugging and unplugging, we observe in 4 that the smartphone was also highly relevant, denoting, that the subjects performing that activity first performed a body movement, such as walking to the light switch or bending toward a power plug. Recognizing simple pressing and pulling actions is relevant for people who are bedridden due to illness, injury, or some other physical condition.

Raising or lowering actions (group three) requires the subject to be more mobile in his arms, enabling movements such as putting or taking an object from the floor or a high position - see Figure 5 and contains 13 such activities. The highest mean F1 scores of the activities in this group is reached by training all the devices together (74.94 %) followed by the smartwatch training F1 score (66.04 %) as shown in Table 2 Such actions are prevalent in physical exercises and could be recognized to check if a person complies with the recommended list of exercises at home.

Body Movements (group four) contains nine more complex exercises where the whole body is involved as shown in Figure 6. Being able to wear or take off a jacket on its own requires complex mobility in the upper body. Transitions such as sitting down and getting up are already part of functional assessments such as the Timed-Up-and-Go (TUG). The TUG is a common geriatric assessment test that can be recognized using wearables and used for motor symptoms assessment in subjects with Parkinson's Disease (Kleiner et al., 2018). Our results show, that a smartphone in the subject's left pocket can also recognize these activities without problems, achieving an F1 score almost as high as the smartwatch (72.94 % and 74.23 %) - see Table 2.

The activities in group five are four movements where the patient raises his arm to their head. In this case, the smartwatch achieves the highest performance (79.23 %) - see Table 2. In Figure 7 we see that the smart glasses are also highly relevant in measuring the movement of the head, for example, while eating, gargling and talking by telephone.

The group of continuous Hand or Head Movements (group six) is different from the other groups because its nine activities are characterized by repetitive similar movements - see 8. Performing repetitive movements is also associated with physical exercises for example in monitoring the evolution of chronic diseases.

In our study we observed, that the smartwatch individually performed very well for the classification having only a 4% lower F1 score than the F1 score achieved by the training of all the devices at once. We can argue that the difference is negligible in clinical practice and that the sensibility of the smartwatch is high enough when considering that using all the devices instead of one, comes with many more challenges.

We want to point out that our contribution is essential for future studies and lies in offering a comprehensive overview of activities that are potentially relevant for HAR in the context of health-related research. Additionally, we have shown which activities can be recognized well by which specific sensor, which enables future studies to choose the specific sensor that works best for their intended research.

6 CONCLUSION

We have trained and evaluated the DeepConvLSTM model for the classification of a large number of diverse activities collected from three devices and compared the classification results with those from benchmark models, such as CNN and LSTM. The placement of the devices on different body parts highlights the strengths of each sensor and its practical significance for future research.

Combining the sensors might improve the accuracy of HAR, but there are some cases, especially in the health-related field where it is not realistic to wear all three devices at once. At the same time, training models with multiple sensors is challenging and requires the synchronisation of the data and specialised models that can deal with different sampling frequencies. We have shown, that the DeepConvLSTM is able to overcome these challenges and outperform other models such as CNN and LSTM to classify 53 different activities.

This diversity of the dataset we used, containing the largest number of classes, is important in facilitating an in-depth analysis of the performance of HAR using three different devices. Futhermore, the comprehensive grouping and analysis of activities in six distinct categories contribute to the relevance of our study for health-related assessments.

Together, these contributions advance the understanding of HAR, especially in the context of health monitoring and assessment.

ACKNOWLEDGEMENTS

The study is funded by the German Federal Ministry of Education and Research (Project No. 01ZZ2007).

REFERENCES

Abbaspour, S., Fotouhi, F., Sedaghatbaf, A., Fotouhi, H., Vahabi, M., and Linden, M. (2020). A comparative analysis of hybrid deep learning models for human activity recognition. *Sensors*, 20(19):5707.

Aboo, A. K. and Ibrahim, L. M. (2022). Human activity recognition using a hybrid lstm-cnn deep neural network. *Webology*.

Augustinov, G., Nisar, M. A., Li, F., Tabatabaei, A., Grzegorzek, M., Sohrabi, K., and Fudickar, S. (2023). Transformer-based recognition of activities of daily living from wearable sensor data. In *iWOAR '22*, pages 1–8, Rostock, Germany. ACM.

Bock, M., Hoelzemann, A., Moeller, M., and Laerhoven, K. V. (2022). Investigating (re)current state-of-the-art in human activity recognition datasets. *Frontiers in Computer Science*, 4.

Ciortuz, G., Grzegorzek, M., and Fudickar, S. (2023). Effects of time-series data pre-processing on the transformer-based classification of activities from smart glasses. In *iWOAR '23*, New York, NY, USA. Association for Computing Machinery.

Dave, R., Seliya, N., Vanamala, M., and Tee, W. (2022). Human activity recognition models using limited consumer device sensors and machine learning. *CoRR*, abs/2201.08565.

Friedrich, B., Cauchi, B., Hein, A., and Fudickar, S. (2019). Transportation mode classification from smartphone sensors via a long-short-term-memory network. page 709–713, New York, NY, USA. Association for Computing Machinery.

Gil-Martín, M., San-Segundo, R., Fernández-Martínez, F., and de Córdoba, R. (2020). Human activity recognition adapted to the type of movement. *Computers & Electrical Engineering*, 88:106822.

Hans Van Remoortel, S. G., Raste, Y., Burtin, C., Louvaris, Z., Gimeno-Santos, E., Langer, D., Glendenning, G., Hopkinson, N., Vogiatzis, I., Peterson, B., Wilson, F., Mann, B., Rabinovich, R., Puhan, M., and Troosters, T. (2012). Validity of activity monitors in health and chronic disease: A systematic review. *The international journal of behavioral nutrition and physical activity*, 9:84.

Hellmers, S., Kromke, T., Dasenbrock, L., Heinks, A., Bauer, J. M., Hein, A., and Fudickar, S. (2018). Stair climb power measurements via inertial measurement units - towards an unsupervised assessment of strength in domestic environments. In *BIOSTEC 2018 - HEALTHINF*, pages 39–47, Funchal.

Hussain, Z., Sheng, Q. Z., and Zhang, W. E. (2020). A review and categorization of techniques on device-free human activity recognition. *Journal of Network and Computer Applications*, 167:102738.

Irshad, M. T., Nisar, M. A., Huang, X., Hartz, J., Flak, O., Li, F., Gouverneur, P., Piet, A., Oltmanns, K. M., and Grzegorzek, M. (2022). Sensehunger: Machine learning approach to hunger detection using wearable sensors. *Sensors*, 22(20).

Kleiner, A. F. R., Pacifici, I., Vagnini, A., Camerota, F., Celletti, C., Stocchi, F., De Pandis, M. F., and Galli, M. (2018). Timed up and go evaluation with wearable devices: Validation in parkinson's disease. *Journal of Bodywork and Movement Therapies*, 22(2):390–395.

Li, F., Shirahama, K., Nisar, M. A., Huang, X., and Grzegorzek, M. (2020). Deep transfer learning for time series data based on sensor modality classification. *Sensors*, 20(15):4271.

Li, F., Shirahama, K., Nisar, M. A., Köping, L., and Grzegorzek, M. (2018). Comparison of feature learning methods for human activity recognition using wearable sensors. *Sensors*, 18(2).

Mahmud, S., Tonmoy, T. H. M., Bhaumik, K. K., Rahman, M. A. K., Amin, A. M., Shoyaib, M., Khan, M. A. H., and Ali, A. A. (2020). Human activity recognition from wearable sensor data using self-attention.

Mekruksavanich, S., Jitpattanakul, A., Youplao, P., and Yupapin, P. (2020). Enhanced hand-oriented activity recognition based on smartwatch sensor data using lstms. *Symmetry*, 12:1570.

Morshed, M. G., Sultana, T., Alam, A., and Lee, Y.-K. (2023). Human action recognition: A taxonomy-based survey, updates, and opportunities. *Sensors*, 23(4).

Nisar, M. A., Shirahama, K., Li, F., Huang, X., and Grzegorzek, M. (2020). Rank pooling approach for wearable sensor-based ADLs recognition. *Sensors*, 20(12):3463.

Ordóñez, F. J. and Roggen, D. (2016). Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1).

Phyo, P. and Byun, Y. (2021). Hybrid ensemble deep learning-based approach for time series energy prediction. *Symmetry*, 13:1942.

Roggen, D., Calatroni, A., Nguyen-Dinh, L.-V., Chavarriaga, R., and Sagha, H. (2012). OPPORTUNITY Activity Recognition. UCI Machine Learning Repository.

Roy, B., Malviya, L., Kumar, R., Mal, S., Kumar, A., Bhowmik, T., and Hu, J. (2023). Hybrid deep learning approach for stress detection using decomposed eeg signals. *Diagnostics*, 13.

Uddin, M. Z. and Soylu, A. (2021). Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning. *Scientific Reports*, 11(1).

Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., and Alshurafa, N. (2022). Deep learning in human activity recognition with wearable sensors: A review on advances. *Sensors*, 22(4).