Individual Convolution of Ankle, Hip, and Wrist Data for Activities-of-Daily-Living Classification

Lee B. Hinkle
Department of Computer Science
Texas State University
San Marcos, TX
leebhinkle@txstate.edu

Vangelis Metsis
Department of Computer Science
Texas State University
San Marcos, TX
vmetsis@txstate.edu

***Abstract*—The Activities of Daily Living (ADL) include activities such as brushing teeth, sweeping, and walking that are critical to on-going health, especially in older adults. Activities may be determined using recorded video and 2D-CNNs, however video recordings present privacy and coverage challenges in personal spaces. Smartphones and newer wristworn devices that record motion data can also be used for activity recognition tasks. Ankle or shoe-based devices such as the retired Nike+ sensor are less common, however ear-based devices which may record head movement are gaining popularity. In this work we use accelerometer data from a recently released dataset using devices placed on the ankle, hip, and wrist. First, we evaluate a simple 1D-CNNs ability to classify the 17 included activities in subject-dependent and subject-independent analysis. Then we process the accelerometer data from the three sensors individually to evaluate each location's ability to predict activities. Finally, we develop a functional model which independently executes a 1D-CNN for each sensor's data and combines the results using Global Average Pooling. The functional model achieves a *subject-independent* accuracy of 70.7%.**

***Index Terms*—human activity recognition, wearable devices, wristband, smartphone, classification labels, visualization**

I. INTRODUCTION

Activities of Daily Living (ADL) have been a part of overall health assessment for decades [1]. Measuring an individual's ability to carry out common tasks that enable independence can have predictive benefits regarding overall quality of life and care requirements. Traditional and newer deep-learning models have been shown to be capable of Human Activity Recognition (HAR) which is a subset of ADLs generally limited to exercise and mobility classes such as walking, running, sitting, and standing. Example HAR datasets using smartphone accelerometer data include MobiAct [2], UCI HAR [3], and UniMiB SHAR [4]. In the literature and published datasets, there are a varied number of activities included in each category. Datasets that include more than movement and exercise physical activity are typically deemed ADL, frequently with an accompanying set of fall data. Activities included in medical studies which rely on practitioner diagnosis (versus device-based measurement) often include activities which are critical to health and wellness but have privacy concerns which must be carefully addressed, e.g., toileting. The remainder of this work relies on the publicly available dataset published by Leotta et al. [5] containing data of daily living activities recorded with wearable devices and has been used in other works [6].

For this dataset, the data were collected using two different types of devices in three locations. Actigraph GT9X Link devices were used for hip and ankle locations. These devices incorporate 9-axis sensors measuring the x,y,z components using an accelerometer, a magnetometer (relation to the earth's magnetic field), and a gyroscope (angular rotation - pitch, roll, and yaw). The hip and ankle data were collected at a sample rate of 100 samples per second. An Actigraph Centrepoint device was worn on the subject's dominant wrist for wrist data collection. This device only records accelerometer data, it does not include a magnetometer or gyroscope, however, it has a higher sample rate of 256 samples per second.

Given the highly similar activities in this dataset we summarize the key contributions of this work as:

1) Evaluation of the delta between the higher subject-dependent and lower subject-independent results using a 1D-CNN.
2) Evaluation of a 1D-CNN's ability to predict the activity using the ankle, hip, or wrist accelerometer sensor data individually.
3) A functional 1D-CNN model which performs convolution on each sensor's data independently in a multi-modal fashion and pools the results to determine the final activity prediction.

The remainder of the paper is organized as follows: Section II describes the processing of the data for input into the classifiers and the methodology used to separate the subjects into the training, validation, and test groups to conduct a subject-independent analysis. Section III describes the sequential model including the contribution of each sensor and the improvement in a subject-dependent scheme. Section IV presents the classification report for the improved functional model which independently executes convolutions on each sensor and pools the results. The model's performance for each subject is shown (hold-one-subject out). We have found this data to be informative regarding subject-to-subject variation. Section V summarizes our findings and future work.

978-1-6654-6934-0/22/$31.00 ©2022 IEEE

II. SUBJECT ALLOCATION AND DATA PROCESSING

When training models involving multiple subjects, the separation into training, validation, and test groups can have a significant impact on the accuracy. Typically, time-series data are placed into sliding windows of several seconds each with or without overlap. Each window is assigned a label based on the mode activity occurring during that time. Since there are many windows contributed from each subject a single subject's data may be present in more than one group. The resulting subject-dependent results will be higher than if the test subject's data is kept separate during model training and hyper-parameter tuning. For our research we are interested classifying previously unseen subjects so our focus is subject-independent analysis.

For the subject-independent analysis, the eight male subjects were split with four subjects allocated to training, two allocated to validation, and the remaining two used for the test group. The subjects were sorted by height and allocated among the three groups ensuring there was a mixture of heights in each group. Unfortunately, there are only two subjects who are left-hand dominant so it was not possible to place a left-handed subject in each of the three groups. The validation accuracy was measured using the two validation subjects with the training completed on the four subjects in the training group. The final subject allocation is shown in Table I. The test subject data was not used to tune the model or hyper-parameters to maintain the subject independence.

TABLE I
THE EIGHT SUBJECTS WERE DIVIDED AMONG THE TRAINING, VALIDATION, AND TEST GROUPS BY SORTING ON HEIGHT AND ALLOCATING TO ACHIEVE AN AVERAGE IN EACH GROUP OF APPROXIMATELY 179CM.

Subject	Age	Height (cm)	Weight (kg)	Dom. Hand	Group
1	37	183	81	Right	Training
2	25	172	53	Right	Training
3	29	176	60	Right	Training
4	27	183	72	Right	Validation
5	28	173	52	Right	Test
6	23	184	90	Right	Test
7	25	186	76	Left	Training
8	27	175	76	Left	Validation

Since the wrist sensor was limited to acceleration data, only the acceleration data were used for the hip and ankle sensors as well. The ankle and hip gyroscope and magnetometer data present in the published dataset were not used. For each sample, the magnitude of the acceleration vector was computed using the individual component accelerations, this yields the total acceleration independent of orientation. The constant 1g acceleration due to gravity (9.8m/s^2) was subtracted so the remaining value represents the magnitude of acceleration due to movement. The wrist data were down-sampled to 100Hz to match the ankle and hip sample rate by placing the data into a DateTime indexed Pandas dataframe and invoking the `resample` method. Finally the data were divided using a sliding-window of 300 samples representing 3 seconds of time with no overlap.

III. SEQUENTIAL MODEL AND RESULTS

The sequential model used is a straight-forward one: two convolution layers extract features from the time-series accelerometer data, followed by dropout and max-pooling layers to prevent over-fitting, then a flatten layer passes to a fully connected layer, with a final fully connected layer to predict the probability for each of the 17 classes.

The subject-dependent results where the data from each subject is present in the training, validation, and test sets when allocated using non-group-based stratification shows the higher accuracy of 80.4% than can be expected using subject based allocation. Notably, the 'relax' category was classified correctly 56/63 times (89%) which is much different than the best-case subject-independent results.

The subject-independent allocation was also run using the component acceleration in the x-axis, y-axis, and z-axis. The accuracy was significantly lower and the training loss curves showed clear signs of over-fitting with the validation loss increasing for each subsequent epoch. This is consistent with our experience that while component acceleration can yield higher accuracies in subject-dependent cases it does not generalize well in subject-independent cases.

Data from each of the sensors were also run individually to assess their stand-alone ability to predict the activity. The highest prediction rate was for the wrist sensor followed by the hip sensor and finally the ankle sensor. This is consistent with literature [7] and we also found it to be consistent within a range of model parameters. In all three cases the accuracy was less than when the three sensors were used together. The results are summarized in Table II.

TABLE II
SUBJECT-DEPENDENT ACCURACY USING TOTAL ACCELERATION, FOLLOWED BY SUBJECT-INDEPENDENT ACCURACIES USING: TOTAL ACCELERATION, COMPONENT ACCELERATION, AND TOTAL ACCELERATION FOR EACH INDIVIDUAL SENSOR. ALSO LISTED IS THE HIGHEST MIS-CLASSIFIED ACTIVITY.

Setup	Accuracy	Most Mis-Classified (# instances)
Sub Dep(3*$\|\|accel\|\|$)	80.4%	sweeping as vacuuming (12)
Sub Ind(3*$\|\|accel\|\|$)	63.1%	relax as laptop (78)
Sub Ind(3*$accel_{x,y,z}$)	38.5%	facewashing as vacuuming (59)
Sub Ind(Wrist$\|\|accel\|\|$)	52.7%	relax as laptop (78)
Sub Ind(Hip$\|\|accel\|\|$)	38.4%	keyboard writing as relax (79)
Sub Ind(Ankle$\|\|accel\|\|$)	29.1%	teethbrush as face washing (71)

IV. FUNCTIONAL MODEL AND RESULTS

For the functional model each sensor's data are processed using separate 1D-CNNs. The result from each of the three convolutions is then concatenated and input into a Global Average Pooling layer [8]. The full model is shown in Figure 1. This model achieves a subject-independent accuracy of 70.7% which may seem low but this is for 17 classes and truly subject independent - the test subjects were not used for tuning or training. This is an improvement of 7.6 percentage points over the highest subject-independent accuracy of 63.1% reported in Table II. The classification report is shown in Table III and the resulting confusion matrix is shown in Figure 2.

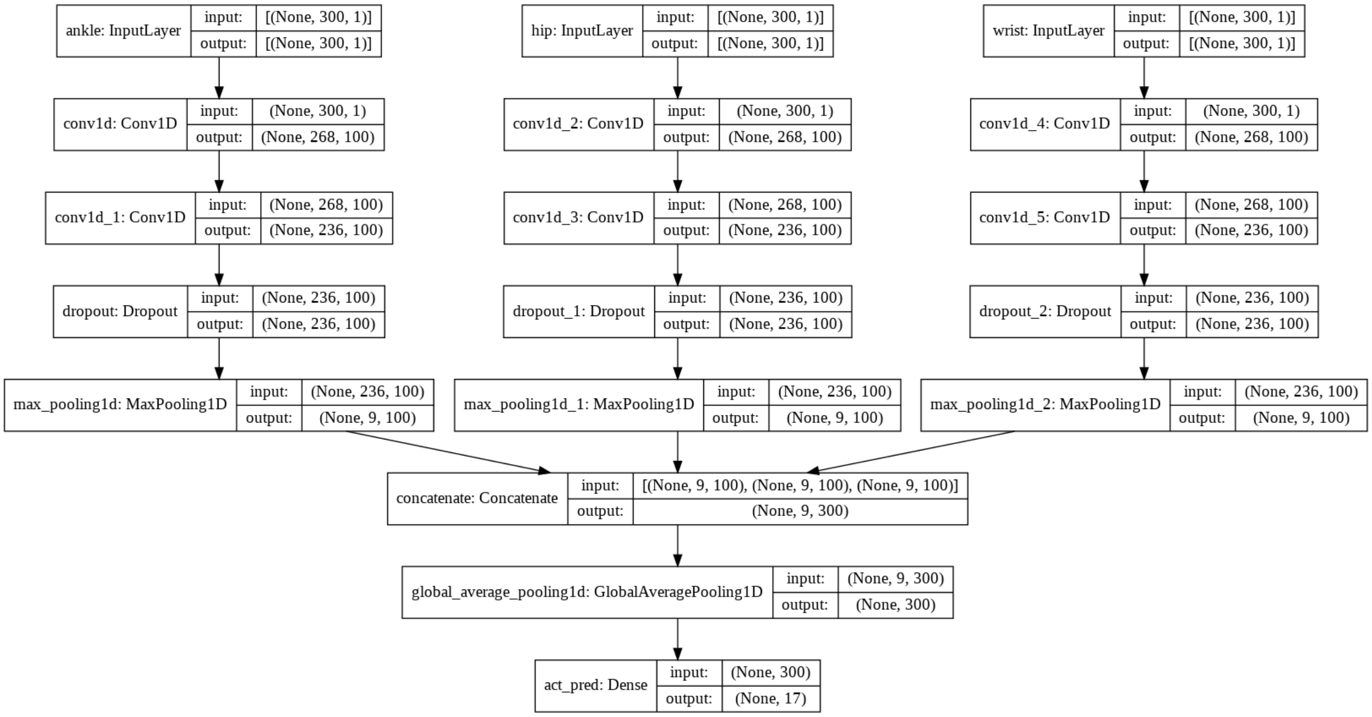


Fig. 1. The final Keras functional model, 1D-CNNs are used on each of the ankle, hip, and wrist data with a final GlobalAveragePooling layer operating on the concatenated data.

TABLE III
CLASSIFICATION REPORT FOR THE FUNCTIONAL MODEL USING THE SUBJECT-INDEPENDENT TEST SET. NOTE THE LOW F1-SCORES FOR RELAX AND LAPTOP.

	precision	recall	f1-score	support
RELAX	0.00	0.00	0.00	78
KEYBOARD_WRITING	0.58	0.97	0.73	79
LAPTOP	0.18	0.19	0.18	79
HANDWRITING	0.98	0.78	0.87	80
HANDWASHING	0.66	0.49	0.56	78
FACEWASHING	0.75	0.72	0.74	79
TEETHBRUSH	0.51	0.62	0.56	78
SWEEPING	0.76	0.68	0.72	79
VACUUMING	0.75	0.76	0.75	78
EATING	0.86	0.94	0.90	78
DUSTING	0.82	0.95	0.88	79
RUBBING	0.99	0.85	0.91	80
DOWNSTAIRS	1.00	0.77	0.87	39
WALKING	0.88	0.83	0.85	77
WALKING_FAST	0.82	0.92	0.87	76
UPSTAIRS_FAST	0.72	1.00	0.84	18
UPSTAIRS	1.00	0.97	0.98	32
accuracy			0.71	1187
macro avg	0.72	0.73	0.72	1187
weighted avg	0.70	0.71	0.70	1187

A final evaluation was performed using the functional model and hold-one-subject out. As shown in Figure 3 there is significant difference in the mean and standard deviation of the accuracy obtained for each of the subjects.

V. CONCLUSION AND FUTURE WORK

Consistent with the original dataset [5] providers we find that of the three sensor locations, the wrist data provides greater information versus the ankle and hip sensors. We have confirmed that when using deep learning models the *subject-dependent* accuracies are significantly better than the *subject-independent* accuracies which reflect the expected performance for a previously unseen subject. Total acceleration reduces over-fitting and provides higher overall accuracies versus component acceleration. Our functional model which performs convolution on each sensor location independently produces the highest *subject-independent* accuracy of 70.7% for the 17-class problem.

The remaining mis-classifications are among activities with obvious similarities: relax-laptop-handwriting, handwashing-facewashing-teethbrush, and vacuuming-sweeping. Future work includes the refinement of each sensor's data processing and hyper-parameter tuning to enable greater separation among the slight differences in motion in these activities. Additionally, the rotation of the wrist is a potential source of arm motion activity; this would require new data collection using a small wrist sensor that includes a gyroscope in addition to the accelerometer.

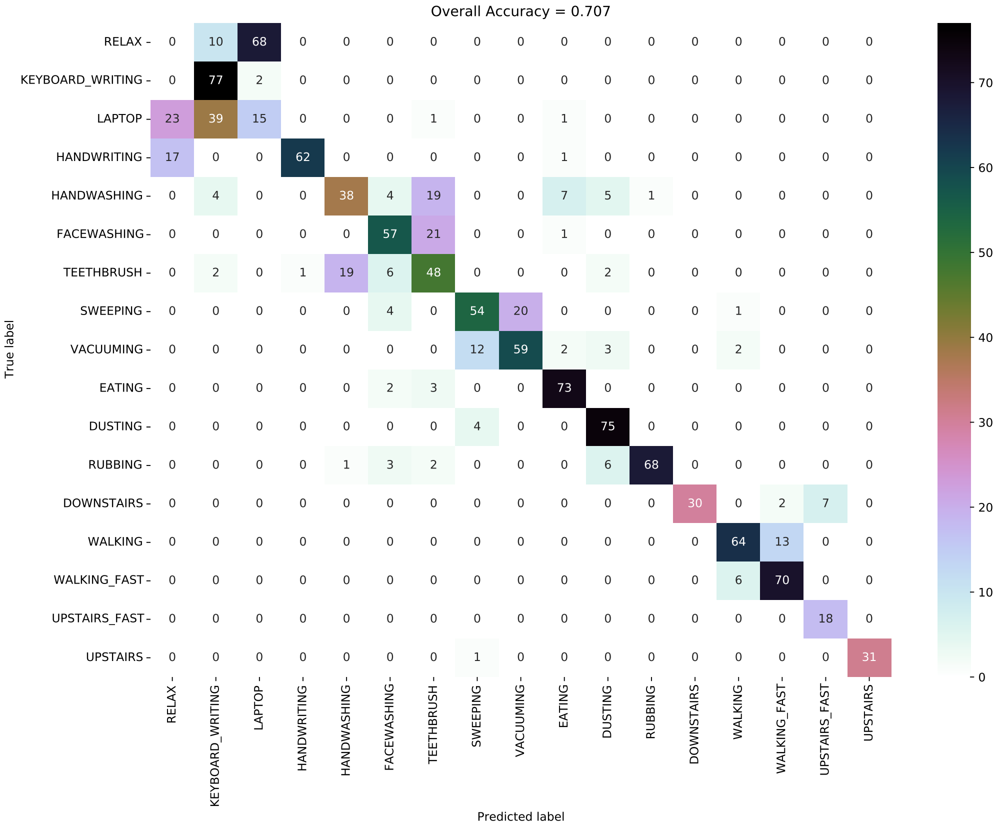
The source code for this work is available on our IMICS Lab github repository.[1]

ACKNOWLEDGMENT

Our thanks and gratitude to the researchers who have collected, organized, and published HAR and ADL datasets. We also appreciate the work of the teams responsible for scikit-learn, Matplotlib, seaborn, and Keras.

---

[1] https://github.com/imics-lab/Indiv-CNN-Ankle-Hip-Wrist



True label	RELAX	KEYBOARD_WRITING	LAPTOP	HANDWRITING	HANDWASHING	FACEWASHING	TEETHBRUSH	SWEEPING	VACUUMING	EATING	DUSTING	RUBBING	DOWNSTAIRS	WALKING	WALKING_FAST	UPSTAIRS_FAST	UPSTAIRS
RELAX	0	10	68	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KEYBOARD_WRITING	0	77	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LAPTOP	23	39	15	0	0	0	1	0	0	1	0	0	0	0	0	0	0
HANDWRITING	17	0	0	62	0	0	0	0	0	1	0	0	0	0	0	0	0
HANDWASHING	0	4	0	0	38	4	19	0	0	7	5	1	0	0	0	0	0
FACEWASHING	0	0	0	0	0	57	21	0	0	1	0	0	0	0	0	0	0
TEETHBRUSH	0	2	0	1	19	6	48	0	0	0	2	0	0	0	0	0	0
SWEEPING	0	0	0	0	0	4	0	54	20	0	0	0	0	1	0	0	0
VACUUMING	0	0	0	0	0	0	0	12	59	2	3	0	0	2	0	0	0
EATING	0	0	0	0	0	2	3	0	0	73	0	0	0	0	0	0	0
DUSTING	0	0	0	0	0	0	0	4	0	0	75	0	0	0	0	0	0
RUBBING	0	0	0	0	1	3	2	0	0	0	6	68	0	0	0	0	0
DOWNSTAIRS	0	0	0	0	0	0	0	0	0	0	0	0	30	0	2	7	0
WALKING	0	0	0	0	0	0	0	0	0	0	0	0	0	64	13	0	0
WALKING_FAST	0	0	0	0	0	0	0	0	0	0	0	0	0	6	70	0	0
UPSTAIRS_FAST	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0
UPSTAIRS	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	31

Fig. 2. Confusion Matrix using the final model with ankle, hip, and wrist data input into individual 1D-CNNs followed by a Global Average Pooling layer.

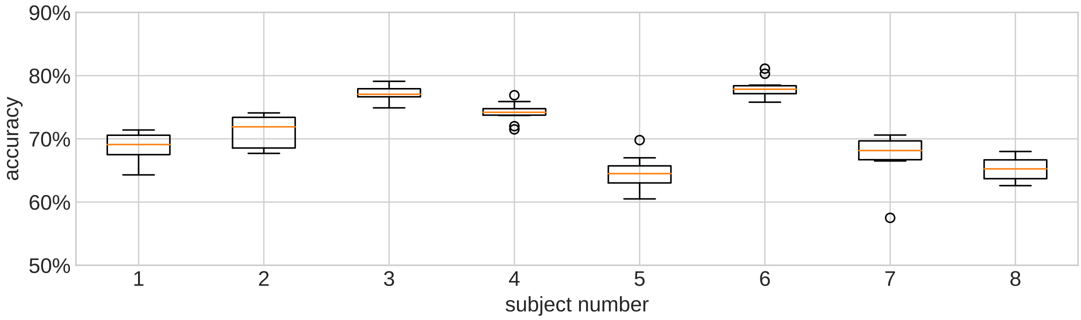


Fig. 3. A box plot showing the mean accuracy over ten runs of the final model for each subject using hold-one-subject out. In each of the eight cases the model was trained using data from the other seven subjects.

REFERENCES

[1] J. M. Wiener, R. J. Hanley, R. Clark, and J. F. Van Nostrand, "Measuring the activities of daily living: Comparisons across national surveys," *Journal of gerontology*, vol. 45, no. 6, pp. S229–S237, 1990.

[2] G. Vavoulas, C. Chatzaki, T. Malliotakis, M. Pediaditis, and M. Tsiknakis, "The mobiact dataset: Recognition of activities of daily living using smartphones." in *ICT4AgeingWell*, 2016, pp. 143–151.

[3] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones." in *Esann*, vol. 3, 2013, p. 3.

[4] D. Micucci, M. Mobilio, and P. Napoletano, "Unimib shar: A dataset for human activity recognition using acceleration data from smartphones," *Applied Sciences*, vol. 7, no. 10, 2017. [Online]. Available: https://www.mdpi.com/2076-3417/7/10/1101

[5] M. Leotta, A. Fasciglione, and A. Verri, "Daily living activity recognition using wearable devices: A features-rich dataset and a novel approach," in *Pattern Recognition. ICPR International Workshops and Challenges*, A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, and R. Vezzani, Eds. Cham: Springer International Publishing, 2021, pp. 171–187.

[6] A. Fasciglione, M. Leotta, and A. Verri, "Improving activity recognition while reducing misclassification of unknown activities," in *2021 IEEE 30th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, 2021, pp. 153–158.

[7] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "How accurately can your wrist device recognize daily activities and detect falls?" *Sensors*, vol. 16, no. 6, p. 800, 2016.

[8] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.