2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

Activity Recognition in Parkinson's Patients from Motion Data Using a CNN Model Trained by Healthy Subjects

Shelly Davidashvilly[1], Murtadha Hssayeni[1], Christopher Chi[1], Joohi Jimenez-Shahed[2], and Behnaz Ghoraani[1, *]

Abstract— **Physical activity recognition in patients with Parkinson's Disease (PwPD) is challenging due to the lack of large-enough and good quality motion data for PwPD. A common approach to this obstacle involves the use of models trained on better quality data from healthy patients. Models can struggle to generalize across these domains due to motor complications affecting the movement patterns in PwPD and differences in sensor axes orientations between data. In this paper, we investigated the generalizability of a deep convolutional neural network (CNN) model trained on a young, healthy population to PD, and the role of data augmentation on alleviating sensor position variability. We used two publicly available healthy datasets - PAMAP2 and MHEALTH. Both datasets had sensor placements on the chest, wrist, and ankle with 9 and 10 subjects, respectively. A private PD dataset was utilized as well. The proposed CNN model was trained on PAMAP2 in k-fold cross-validation based on the number of subjects, with and without data augmentation, and tested directly on MHEALTH and PD data. Without data augmentation, the trained model resulted in 48.16% accuracy on MHEALTH and 0% on the PD data when directly applied with no model adaptation techniques. With data augmentation, the accuracies improved to 87.43% and 44.78%, respectively, indicating that the method compensated for the potential sensor placement variations between data.**

Clinical Relevance— **Wearable sensors and machine learning can provide important information about the activity level of PwPD. This information can be used by the treating physician to make appropriate clinical interventions such as rehabilitation to improve quality of life.**

I. Introduction

The field of human activity recognition research with wearable body sensors is notorious for data with high variability and small quantities. The challenges are exacerbated with data from elderly people and more specifically the Parkinson's population, which are subject to larger intra-class variability and noisier labels [1]. Since it is difficult to collect generous amounts of data from the Parkinson's population, or any population for that matter, leveraging knowledge from similar domains is a common approach. In doing so, models tend to struggle to generalize across domains, largely due to differences in the axis orientation between sensors (caused by different placements of sensors on the body). Augmenting data in such a way where models become indifferent to variations in sensor orientations can help the challenge of generalizing to different domains, notably from younger, healthier subjects, to patients with Parkinson's disease (PwPD). However, the existing research has paid less attention to generalizability of deep models and data augmentation in disease population. To address this shortcoming, in this paper, we designed a convolutional neural network (CNN) model and investigated the generalizability of the model when training on a young, healthy population and testing on PwPD. Also, we implemented standard data augmentation methods after matching the sensors' orientation between the different datasets.

II. Related work

Straczkiewicz *et al.* [2] evaluated the effect of sensor placement variations on physical activity classification from a study on 45 older adults in a 7-day collection of free-living data. Such environments, which are more common in applications of wearable body sensors, leave researchers with less ability to enforce a sensor-wearing protocol. They observed that patients tended to deviate from said protocol when affixing sensors upon themselves in their free-living setting, which led to a significant impact in activity recognition error. The authors in [3] used the rotation, permutation, and time-warping data augmentation methods, coupled with transfer learning to utilize knowledge gained from the younger populations to older adults. They found that rotation and permutation were among the most successful data augmentation methods. Um *et al.* [1] utilized the same augmentation techniques for detecting medication states in PwPD. The best performing methods from their experiments were rotation, permutation, and time-warping. These results indicated that the largest sources of variability were 'different sensor placements between participants and event locations in an arbitrarily segmented window. These papers typically explore the techniques to improve model generalizability within the same dataset or with transfer learning. The experiments presented throughout the paper are novel as they explore the impact of signal augmentations on a proposed CNN model's ability to generalize from healthy population to PwPD directly without transfer learning.

III. METHODOLOGY

A. Datasets

The experiments utilized two publicly available datasets, MHEALTH [4, 5] and PAMAP2 [6, 7], and a private dataset, referred to as the PD (Parkinson's Disease) dataset [9].

- MHEALTH: Positioned sensors on the chest, right wrist, and left ankle of 10 subjects performing 12 activities, with a sampling frequency of 50 Hz and acceleration unit of m/s^2.
- PAMAP2: Sensors were placed on the dominant-side wrist and ankle for 9 subjects performing 18 activities. Subject 8 was the only participant with the wrist and ankle sensors placed on the left side-body. The sampling

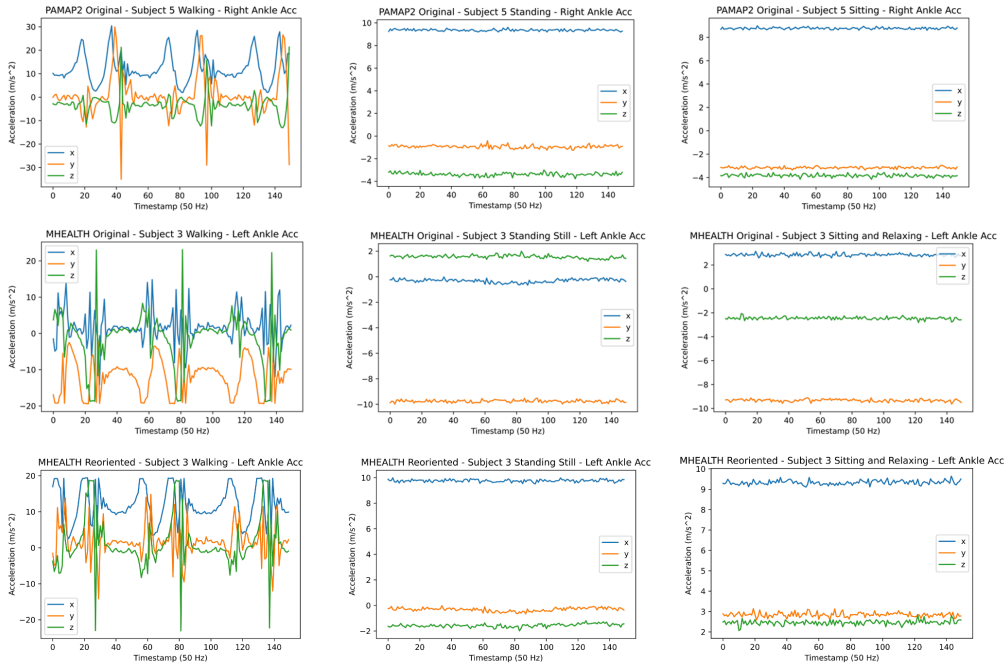
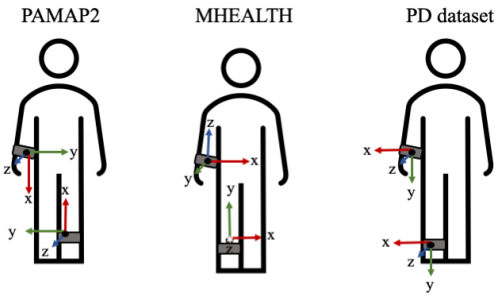


Figure 1. Sample activity plots from PAMAP2, MHEALTH, and MHEALTH after reorienting the axes to match PAMAP2.

frequency was 100 Hz and the accelerometer units were the same as MHEALTH.

- PD: Collected from 14 subjects with idiopathic Parkinson's Disease under approval of the institutional review boards of University of Rochester and Great Lakes NeuroTechnologies. Sensors were mounted on the wrist and ankle of the patient's most affected side as they performed four rounds of daily living activities. Subjects stopped their medication the night prior to data collection to allow the sensors to capture their dyskinesia. The sampling frequency was 128Hz and the acceleration units were standard gravity.



	Axes reorientation of MHEALTH	Axes reorientation of PD dataset
WRIST	X -> Y Y -> Z -Z -> X	-X -> Y Y -> X Z -> Z
ANKLE	-X -> Y Y -> X -Z -> Z	X -> Y -Y -> X Z -> Z

Figure 2. Sensors' placement in all the datasets and the manual orientation required to align the sensor orientation of MHEALTH and PD to PAMAP2.

Figure 1 shows sample signals in two healthy datasets for different activities before and after reorientation. Reoriented MHEALTH signals in the last row of Figure 1 show a similar pattern to PAMAP2 signals in the first row. Figure 2 provides the sensor placement information.

B. Data Preprocessing

Only accelerometer data from wrist and ankle sensors were used to match the PD dataset. Gyroscope data was removed, since MHEALTH had poor resolution quality for this data. Doing so improved the results for cross testing across datasets. Any data labeled as 'null class' - referring to motions that were not characterized by specified activities - were removed. PAMAP2 had some missing values, which were removed. The PAMAP2 and PD datasets were down sampled from 100Hz and 128Hz, respectively, to match the MHEALTH frequency of 50Hz. A generic normalization was applied to the data prior to segmentation by finding the mean and STD of each axis in the training signals and then subtracting the mean and dividing by the STD from the training and testing data. The data was segmented into windows of 150 timesteps with 50% overlap.

5 seconds of data were deleted from the beginning and end of each labeled activity from MHEALTH and PAMAP2 to remove transitions between activities, as suggested by [8]. The 7 common activities between MHEALTH and PAMAP2 were kept for model training and testing (standing, sitting, lying, walking, climbing stairs, cycling, and running). Subject 9 from PAMAP2 did not perform any of the 7 activities and was consequently removed from the analysis. The PD dataset contained the following activities: ambulation, arms resting, cutting, dressing, drinking, unpacking groceries, hair brushing with left hand, and hair brushing with right hand. For cross testing purposes, the ambulation activity was mapped to walking, arms resting/cutting/drinking were primarily done while seated and were mapped to sitting, dressing/unpacking

groceries were mapped to standing, and the hair brushing activities did not correspond well to either of those 3 and were excluded.

Data Augmentation: The data augmentation methods and code were applied to PAMAP2 as provided by [1]. Rotation, jitter, and scaling augmentations were combined and applied to each segment of the original PAMAP2 data to generate 8 new samples per segment. Rotation arbitrarily rotates the axes of a segment to simulate different sensor placements. The jitter method introduces some Gaussian noise to the data to mimic sensor noise. Scaling multiplies data in a window by a random scalar to simulate multiplicative signal noise. The latter two methods are intended to make the model more robust to noisy signals commonly found in Parkinson's patients.

C. CNN Model Architecture

The proposed model (Figure 3) consists of two 1D convolutional and max pooling layers, followed by a global average pooling and dropout layer before the output layer. The model was trained for 60 epochs with a batch size of 64, ReLU activation, learning rate of 0.001, and Adam optimizer with categorical cross-entropy loss. The output layer utilized a SoftMax activation function. Each pooling layer used filters with width 2 and stride 1. The convolutional layers used 16 and 32 kernels, respectively, each with size 3 and stride 1. L1 and L2 bias and kernel regularization with a rate of 0.001 was applied to each convolutional layer. The final model had 2,103 parameters.

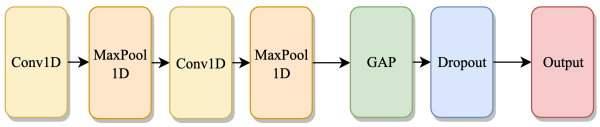


Figure 3. Proposed model architecture

D. Experimental Design

The first set of tests consisted of manually reorienting the axes of one dataset by means of inspecting the graphs of various activity segments visually and deriving a mapping of one dataset's axis orientation to another. The standing, sitting, and walking activities are useful references for generating such a mapping for wrist and ankle sensors, as subjects can generally be expected to have their legs and arms in predictable positions for those activities [1]. Sample graphs of the activities and mappings obtained between the datasets are pictured in Figures 1 and 2. Applying this reorientation mapping on the data requires multiplying columns by -1 if its corresponding axis was negated in the mapping and reordering the columns corresponding to the individual axes accordingly.

The main challenge with this approach lies with the fact that sensor orientations can vary in a multitude of ways across activities and subjects of the same dataset. In these circumstances, a single mapping from one dataset's axes orientation to another is insufficient to capture all the variability. This is observed, for instance, with the wrist sensor data of MHEALTH and subject 8 data from PAMAP2. The wrist sensor graphs for MHEALTH showed more variability in the axis orientations across the three observed activities than the ankle sensors. Similar analysis for subject 8 (left-handed) compared to the other PAMAP2 subjects (right-handed) indicated that the sensor placements on either side-body were potentially the same (e.g., top of wrist and front of ankle for both sides), but the axes orientations presented differently by virtue of the sensors being on opposite sides of the body. Subject 8 consistently showed lower accuracy in the leave-one-subject-out cross validation results that was only fixed with data augmentation. The second set of tests involved augmenting the signal data from PAMAP2, as described in the data preprocessing section.

IV. RESULTS

Two sets of validation were performed. First, models were trained with a leave-one-subject-out cross-validation on PAMAP2 to allow the training, validation, and testing sets to have non-repeated data. Table I reports the average results across the 8 folds for the original and augmented PAMAP2 data. The testing accuracy was 83.98% without data augmentation and 80.74% with data augmentation. The lower accuracy with data augmentation may show the need for a deeper model or reducing the augmentation parameters.

Second, testing on held-out sets (i.e., MHEALTH and PD datasets). For cross-testing, the highest performing model from the source dataset's cross validation was loaded (i.e., PAMAP2) and tested on the target dataset directly. The mean and standard deviation used to normalize the data that trained the source model were applied to the target data. In the cross-testing results recorded in Table II, 'P2O' refers to the PAMAP2 data with its original axes orientation, 'MHR' refers to MHEALTH with its axes orientation matched to PAMAP2, 'PDR' refers to the Parkinson's data with its axes orientation matched to PAMAP2, and P2DA refers to the augmented data of PAMAP2. Without data augmentation, the models performed poorly on healthy and non-healthy populations with 48.16% and 0%, respectively. On the PD dataset, the model was predicting only laying-down label which is not in the activities of PD data. Testing the model with data augmentation on MHEALTH and PD showed significantly higher results than the model trained without data augmentation. With data augmentation, the accuracy was 87.43% on the MHEALTH dataset and 44.78% on the PD dataset. MHEALTH includes the data of healthy subjects, which explains the high accuracy. Also, the PD data did not have activities corresponding exactly to those in the publicly available PAMAP2 and MHEALTH datasets, which could have affected the accuracy performance. For instance, drinking was done seated in the PD data collection and was mapped to sitting in PAMAP2.

TABLE I. CROSS VALIDATION AVERAGE RESULTS

	Accuracy	*Precision*	*Recall*	*F1*
PAMAP2 Original	83.98%	84.59%	83.98%	84.06%
PAMAP2 Data Aug.	80.74%	80.63%	80.74%	78.50%

TABLE II. CROSS TESTING RESULTS

		Accuracy	*Precision*	*Recall*	*F1*
Without Data Aug.	P2O -> MHR	48.16%	61.96%	48.16%	43.95%
	P2O -> PDR	0%	0%	0%	0%
With Data Aug.	P2DA -> MHR	87.43%	89.74%	87.43%	87.33%
	P2DA -> PDR	44.78%	29.22%	44.78%	31.98%
	MH->P2 Transfer Learning [10]	49.30%	X	X	X
	Young-> Healthy LSTM-CNN [11]	~51%	X	X	X

V. DISCUSSION

We referenced the work of Chen *et al.* [10] as a baseline, which applied transfer learning to the chest sensors of the MHEALTH (source) and PAMAP2 (target) datasets with 49.30% accuracy. The results show an accuracy drop of 16% and 27% when testing on PD dataset in comparison with PAMAP2 and MHEALTH datasets, respectively. Sabahat [11] reports a similar drop when testing the elderly population, but with an accuracy of around 50%.

In the leave-one-subject-out cross validation on the original data, subject 8 displayed lower accuracy compared to the right-handed subjects. Many manual reorientation mappings were attempted, none of which were able to alleviate the discrepancy as well as the data augmentation. Although the average of the data augmentation cross validation is slightly lower, the standard deviation between fold results was considerably higher in the original data cross validation, showing that the data augmentation was able to compensate for the intra-subject variability nicely. Subject 8's accuracy increased from 56.06% to 74.98% when it was held out for testing with the data augmentation applied. The manual reorientation approach is more time consuming and error prone than arbitrarily rotating the axes several times to account for a wider range of possible sensor orientations, but it did increase the cross-testing accuracy from PAMAP2 to MHEALTH when combined with data augmentation versus when the augmented data was cross-tested on the original data. Data augmentation was best able to fix the issue of sensor orientation variability, which alleviates the concern of the leave-one-subject-out cross validation method being prone to intra-subject variability. It is, however, considerably more computationally expensive to augment data samples and train models on the resulting, larger dataset.

VI. CONCLUSION

This work investigated the need for transferring models learned on healthy population data to disease population, PD in this case. When it comes to signal data from wearable body sensors, a great portion of what needs to be learned across a new domain is different sensor orientations. Our investigations demonstrated the impact of signal augmentation on such model's generalizability with respect to sensor placement variations when applied to healthy and PwPD datasets. Data augmentation improved a deep model's performance from 48.16% to 87.43% in case of cross testing of healthy-to-healthy dataset and 0% to 44.78% in case of healthy to PD. The lower improvement of the latter cross testing suggests that human activity recognition models trained on healthy population need to be further enhanced when are applied on older population and especially people with PD. Our future work will investigate the effect of transfer leaning or domain adaptation in conjunction with the research done in the paper to boost performance.

ACKNOWLEDGMENT

Data analysis was supported by the NSF 1936586 and 1942669, and NSF REU CNS-1950400. Collection of PD dataset was supported by an NIH SBIR grant to Cleveland Medical Devices (lR43NS071882-01A1; T. Mera, PI).

REFERENCES

[1] T. T. Um, F. M. J. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. Fietzek, and D. Kulić, "Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring using Convolutional Neural Networks," ICM, pp. 1-4, 2017.

[2] M. Straczkiewicz, N. W. Glynn, and J. Harezlak, "On Placement, Location and Orientation of Wrist-Worn Tri-Axial Accelerometers during Free-Living Measurements," MPDI Sensors, 2019.

[3] G. Kalouris, E. I. Zacharaki, and V. Megalooikonomou, "Improving CNN-based activity recognition by data augmentation and transfer learning," pp. 1-7, 2019.

[4] O. Banos, R. Garcia, J. A. Holgado-Terriza, M. Damas, H. Pomares, I. Rojas, A. Saez, and C. Villalonga, "mHealthDroid: A Novel Framework for Agile Development of Mobile Health Applications," *Research Center for Information and Communications Technologies, Proceedings of the 6th International Work-conference on Ambient Assisted Living an Active Aging, IWAAL, December 2-5,* 2014.

[5] O. Banos, C. Villalonga, R. Garcia, A. Saez, M. Damas, J. A. Holgado-Terriza, S. Lee, H. Pomares, and I. Rojas, "Design, implementation and validation of a novel open framework for agile development of mobile health applications," *BioMedical Engineering OnLine (BMC)*, vol. 14, no. S2:S6, pp. 1-20, 2015.

[6] A. Reiss and D. Stricker, "Introducing a New Benchmarked Dataset for Activity Monitoring," *The 16th IEEE International Symposium on Wearable Computers (ISWC)*, 2012.

[7] A. Reiss and D. Stricker, "Creating and Benchmarking a New Dataset for Physical Activity Monitoring," *The 5th Workshop on Affect and Behaviour Related Assistance (ABRA)*, 2012.

[8] A. Reiss, "Personalized Mobile Physical Activity Monitoring for Everyday Life", Technical Uni. of Kaiserslautern, 2014

[9] M. D. Hssayeni, J. Jimenez-Shahed, M. A. Burack, and B. Ghoraani, "Dyskinesia Severity Estimation in Patients with Parkinson's Disease Using Wearable Sensors and A Deep LSTM Network," in *2020 42nd Annual International Conference of the IEEE EMBC*, July 2020.

[10] Y. Chen, J. Wang, M. Huang, and H. Yu, "Cross-position activity recognition with stratified transfer learning," 2019.

[11] Fatima, Sabahat. "Activity Recognition in Older Adults with Training Data from Younger Adults: Preliminary Results on in Vivo Smartwatch Sensor Data." ACM SIGACCESS Conference on Computers and Accessibility, pp. 1-4. 2021.