Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring using Convolutional Neural Networks

Terry T. Um
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada N2L 3G1
terry.t.um@gmail.com

Franz M. J. Pfister
Ludwig-Maximilians-Univ. München
Geschwister-Scholl-Platz 1
Munich, Bavaria, Germany 80539
fmj.pfister@me.com

Daniel Pichler
Technical University of Munich
Arcisstraße 21
Munich, Bavaria, Germany 80333
daniel.pichler@tum.de

Satoshi Endo, Muriel Lang, Sandra Hirche
Technical University of Munich
Barerstr 21
Munich, Bavaria, Germany 80333
{s.endo,muriel.lang,hirche}@tum.de

Urban Fietzek
Schön Klinik München Schwabing
Parzivalplatz 4
Munich, Bavaria, Germany 80804
urban.fietzek@schoen-kliniken.de

Dana Kulić
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada N2L 3G1
dana.kulic@uwaterloo.ca

ABSTRACT

While convolutional neural networks (CNNs) have been successfully applied to many challenging classification applications, they typically require large datasets for training. When the availability of labeled data is limited, data augmentation is a critical preprocessing step for CNNs. However, data augmentation for wearable sensor data has not been deeply investigated yet.

In this paper, various data augmentation methods for wearable sensor data are proposed. The proposed methods and CNNs are applied to the problem of classifying the motor state of Parkinson's Disease (PD) patients, which is challenging due to small dataset size, noisy labels, and large within-class variability. Appropriate augmentation improves the classification performance from 76.7% to 92.0%.

CCS CONCEPTS

•Applied computing → Consumer health; •Computing methodologies → *Supervised learning by classification;*

KEYWORDS

data augmentation, wearable sensor, convolutional neural networks, Parkinson's disease, health monitoring, motor state detection

ACM Reference format:
Terry T. Um, Franz M. J. Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. 2017. Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring using Convolutional Neural Networks . In *Proceedings of International Conference on Multimodal Interaction, Glasgow, Scotland, Nov 2017 (ICMI2017),* 8 pages.
DOI: 10.475/123_4

DOI: 10.475/123_4

1 INTRODUCTION

In recent years, convolutional neural networks (CNNs) have shown excellent performance on classification problems when large-scale labeled datasets are available. The successes of CNNs include static image classification (e.g. [17], [12]), and extend to other nonstatic domains, for example, video [15], speech [26], and sentence [16] classification. However, due to the requirement for large labelled datasets, it is challenging to apply CNNs to problems where only small labelled datasets are available.

Data augmentation leverages limited data by transforming the existing samples to create new ones. A key challenge for data augmentation is to generate new data that maintains the correct label, which typically requires domain knowledge. For example, in image recognition, it is known that scaling, rotating, translating, cropping, jittering, and warping of an image, within a small range, do not change its label, thus, these operations can be used as label-preserving data augmentation methods [9]. However, it is not obvious how to carry out label-preserving augmentation in some domains, e.g., wearable sensor data. For example, scaling of the acceleration data may change their labels because some labels are differentiated by the intensity of motion.

In this paper, we consider the problem of motor state detection of Parkinson's Disease patients from wearable sensor data. Most patients with Parkinson's disease (PD) experience motor fluctuations, which are characterized by phases of *hypo-bradykinesia*, i.e. underscaled and slow movement, and *dyskinesia*, i.e. overflowing spontaneous movement. Dopaminergic treatment can alleviate symptoms of bradykinesia while its over-treatment can cause dyskinesia. Thus, An accurate evaluation of a patient's phenomenology is needed for determining the right dose of medication to alleviate the symptoms of bradykinesia and prevent overtreatment-induced dyskinesia. The current evaluation process relies on patient's self-reports and visual observation by the clinician [22].

To automate the evaluation process based on wearable sensor data, it is necessary to collect a large number of labelled data from various PD patients. However, the collecting and labelling process is difficult and time consuming, thus, the amount of collected data is inevitably small. Moreover, the PD data have a challenge of large variability due to various symptom patterns, irrelevant motion interference, and noisy labels, which will be discussed in detail in Section 3.1. As a consequence, monitoring PD motor states using machine learning techniques is a challenging task due to the large variability and the limited availability of the PD data.

In this paper, this challenging task is successfully tackled using a CNN and data augmentation. The contributions of the paper can be summarized as follows:

- A set of approaches for data augmentation of wearable sensor datasets for CNN-based classification.
- Application to the task of PD motor state classification, using a clinician-labeled dataset of 25 PD patients in daily-living conditions.
- Experimental comparison of various data augmentation methods.

The remainder of the paper is organized as follows: first, a brief survey of wearable sensor-based PD monitoring and data augmentation methods is presented in Section 2. The challenges of PD motor state classification and the proposed data augmentation methods are outlined in Section 3. Comparative experimental results using CNNs with various choices of data augmentation methods are presented in Section 4. Finally, the achievements of the research are summarized in Section 5.

2 RELATED WORK

Several researchers have attempted to automate PD motor state classification using machine learning techniques based on wearable sensor data (e.g. [25], [10], [7]). However, most are limited to standardized motor tasks in clinical settings to suppress the interference due to irrelevant hand motions. For example, [25] estimates the severity of tremor, bradykinesia and dyskinesia with multiple accelerometers using support vector machines (SVMs) [3]. Also, [7] applies a CNN to the classification of bradykinesia present and absent states based on the wearable sensor data collected during several motor tasks.

Although research performed with standardized motor tasks has shown successful bradykinesia and dyskinesia detection performance [24], it is tedious for PD patients to perform the motor tasks in their daily life for monitoring the motor states. Also, the amount of collectable data limited to executing specific motor tasks is much smaller than that what could be available without the motor-task constraint, thus, not suitable for a large-scale PD monitoring [18]. The ability to assess motor state from free-living conditions has been identified as an important need for improved assessment and treatment; but to date there is no fully validated system to monitor clinical features or activities in free-living environments [6].

Deep learning (DL) approaches [19] provide a promising methodology to deal with the large variability of large-scale PD data. In [11], a restricted Boltzmann machine (RBM) was used as pretraining for classifying four PD motor states. RBMs have been used for pretraining also in image recognition [20] and speech recognition [5] problems, however, they are rarely used in recent years because dropout [27] and data augmentation [17] provide simpler ways to regularize DL models. Since dropout can be understood as a subarea of data augmentation by multiplying by noise [29], data augmentation can be considered as the most widely-used regularization method for DL approaches in recent literature.

Data augmentation has become an indispensable preprocessing step for achieving peak performance in DL approaches since [17]. Even when large-scale data are available, data augmentation provides additional performance improvements by providing information about the possible transformations of inputs (e.g. [12]). Because of its practical importance, data augmentation has been investigated with systematic experiments, e.g., for image [2] and speech recognition [4]. However, to the best of our knowledge, standardized data augmentation methods for wearable sensor data have not been systematically investigated yet. The development of data augmentation methods for wearable sensor data would play a key role in extending the powerful performance of DL approaches to the wearables domain.

In this paper, we propose various data augmentation methods that enable the classification of PD motor states from wearable data using CNN and a small-scale dataset.

3 PD MOTOR STATE CLASSIFICATION

3.1 Challenges in PD data

The two frequent PD motor states, bradykinesia and dyskinesia, in daily-living motions are considered in this research. A PD patient in the bradykinesia state exhibits decreased movement speed. Since bradykinesia is difficult to differentiate from voluntary rest, the clinician often asks the patient to perform repetitive movements to assess the level of bradykinesia in the clinical setting. Note that PD patients are not asked to perform any specific motor tasks in this research. Bradykinesia can also be accompanied by tremor, which changes the reading of a wearable sensor significantly. As a result, bradykinesia states with tremor become a major source of mispredictions, in addition to the confusion between bradykinesia and a patient who is voluntarily at rest [22].

Dyskinesia is a drug-induced motor state that is characterized by involuntary muscle movements, similar to tics or chorea. Since dyskinesia appears as sudden accelerations and fluctuating movements, it is easier to detect by a human observer as well as by a wearable sensor than bradykinesia. Unlike bradykinesia, which usually requires specific motor tasks to be detected, dyskinesia can be detected in daily-living motions [24] without specific motor tasks. Note that dyskinesisia is different from tremor in that it shows nonrhythmic and flowing movements, but can be a source of mispredictions during classification by a wearable sensor.

Figure 1 illustrates exemplar one minute data windows of both motor states, from a single accelerometer worn on the wrist of PD patients. Bradykinesia data typically appear as constant signals indicating less movement (first row (a),(b)) while dyskinesia data consist of fluctuating movements (second row (a),(b)). However, there are a significant number of examples that deviate from the stereotypical expression ((c),(d)) that make the classification more challenging.

There are several factors that can cause an apparent disagreement between the observed data pattern and the expert label. First, if the body of the patient indicates, e.g., a dyskinesia state, but the hand which wears the wearable sensor does not move because the patient

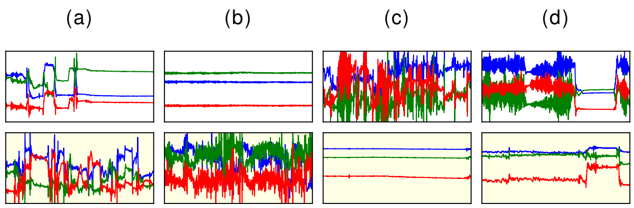

Figure 1: Examples of bradykinesia (white) and dyskinesia (yellow) data in a 1 min window. (a) and (b) show stereotype examples of bradykinesia and dyskinesia while (c) and (d) show the opposite patterns. The blue, red, green represent X,Y,Z signals from the accelerometer, respectively.

is, e.g., holding a chair for suppressing the symptom, the assigned label based on the overall body expression will be mismatched with the recorded data from the wearable device. This kind of voluntary suppression of symptoms happens often in patients' daily life. As a result, the labels assigned under this condition are noisy and make the learning process more challenging.

Another factor that can cause disagreement-labels are changes in the motor state within fixed length windows. For example, the first half of a one minute window may have a different motor state from the second half, as in the first row of Figure 1(d). This problem is caused by arbitrary segmentation into fixed length windows, which may not result in single-labelled windows. This kind of poor segmentation also happens often, especially in the cases where motor states last for less than the fixed length windows.

The last factor that can cause disagreement-labels is the interference of voluntary movements that are irrelevant to evaluating PD motor states. Voluntary movements performed by the patient can make bradykinesia states look like dyskinesia (e.g., waving the hand), and vice versa (e.g., intended slow movements). The voluntary movement interference can be comparable to background interference in image and speech recognition, although the variation of raw data due to the interference is more significant in wearable sensor data with voluntary movements.

While the above issues can be partially solved by a careful and detailed labelling process, it is not easy to realize during data collection from daily-living conditions. The above issues introduce a large within-class variability and significant overlap between the two classes, bradykinesia and dyskinesia. As a result, classification of the two motor states is a challenging task, particularly given a small dataset with noisy labels.

In the next section, various data augmentation methods for addressing the challenges of small and noisily-labelled data are presented.

3.2 Data Augmentation Methods for Wearable Sensor Data

Data augmentation can be considered as an injection of prior knowledge about the invariant properties of the data against certain transformations. Therefore, prior knowledge of the data domain is important for finding the transformations that do not change the data label. If we know the *label-preserving* transformations for the data and the task, additional data can be created by applying the transformations to the raw data. Augmented data can cover unexplored input space, prevent overfitting, and improve the generalization ability of a DL model [9].

In image recognition, it is well-known that minor changes due to jittering (adding noise), scaling, cropping, warping (distorting) and rotating do not change the data labels because theses transformations are likely to happen in real world observations [9]. However, label-preserving transformations for wearable data are not obvious and intuitively recognizable. Moreover, unlike the data augmentation in image recognition, it is difficult to assess the result of data augmentation since the labels of wearable sensor data are not easily determined by eye. Thus, it is necessary to evaluate the effectiveness of various data augmentation methods using systematic comparative experiments.

One factor that can introduce label-invariant variability of wearable sensor data are differences in sensor placement between participants. For example, an upside-down placement of the sensor can invert the sign of the sensor readings. However, different sensor placements do not change the labels, thus, they can be regarded as label-preserving transformations. Therefore, data can be augmented by applying arbitrary *rotations* to the existing data as a way of simulating different sensor placements.

Note that the appropriate choice of data augmentation method can vary depending on the target task. For example, augmentation by rotation may interfere with the learning process of activity recognition if the target task is to classify, for example, pulling up and pulling down motions. Thus, the appropriate transformation for data augmentation should be carefully determined depending on the target task. This is a limitation of data augmentation approaches not only for wearable sensor data, but also for any other data type, e.g., image, speech, etc.

Another factor that can introduce unnecessary variability is the temporal location of activity events, for example, the appearance of tremor. Since the fixed size window segmentation is arbitrary, the location of the observed symptom in the window does not have any meaning. In other words, the same PD data can be represented differently based on the choice of window localization. Thus, we may augment data by perturbing the location of the windows or events.

A potential approach to realize the above augmentation would be to first detect the events, and second, change the temporal location of the event. This problem is similar to object detection (e.g. [8]) or weakly-supervised learning (e.g. [23]) in the image domain in that events or objects may be found anywhere in the viewed space. However, because we have neither the annotations for the locations of the events nor a large amount of data that enables the detection of the events under weak supervision, a simpler approach is needed to perturb the location of the events for this research.

Permutation is a simple way to randomly perturb the temporal location of within-window events. To perturb the location of the data in a single window, we first slice the data into N same-length segments, with N ranging from 1 to 12, and randomly permute the segments to create a new window. This method can create a variety of new data that is made of the slices of the provided data. Although the poor segmentation issue can also arise in this permutation method, the issue is not as significant as in the longer

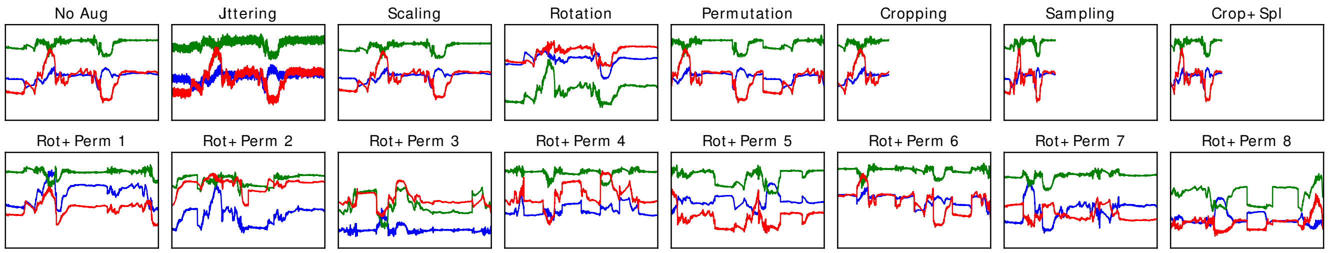

Figure 2: Various data augmentations that are used in the experiments: jittering, scaling, cropping, sampling, rotating, permutating data signals. The bottom row shows several examples generated by the combination of rotation and permutation.

window. Furthermore, permutation can reduce the dependency of the event location, as a result, obtain weakly-invariant features over various locations of the events.

Cropping is another simple approach to diminish the dependency on the event location. By randomly cropping a window into a size M second window, the method reduces the length of the time-series input, thus, decreases the variability due to different event locations. Note that, similar to image cropping, which can inadvertently capture a region of meaningless background, cropping of wearable sensor data can also capture an event-free region, which might change the label. Random cropping along the time axis for every epoch will eventually extract the highest number of possible M second crops from a window, thus converging to a sliding window method with random stride sizes.

Sampling M second samples from a one minute window can be considered as perturbing the data with local translations. That is, irregular intervals by random sampling distort the timesteps between the samples, thus, can introduce variability due to local translations in time. This augmentation method is comparable to warping in image augmentation, although data interpolation after warping is simplified to random sampling in this case.

Jittering and scaling, which are also widely-used in image augmentation, can also be possible candidates for data augmentation for wearable sensor data. Jittering and scaling, can be considered as introducing additive and multiplicative noise, respectively. Since the data gathered from a wearable sensor may contain sensor noise, jittering and scaling of the data as a data augmentation method may increase robustness against noise. Note that jittering introduces element-wise additive noise while scaling introduces window-wise multiplicative noise.

In summaray, jittering, scaling, cropping, warping (sampling) and rotating strategies for image augmentation have been adapted for wearable sensor data augmentation (Figure 2) with a novel permutation augmentation method. In experiments, the performance of PD motor state classification with the proposed data augmentation methods is evaluated with CNNs.

4 EXPERIMENTS

4.1 Data preparation

A dataset of 27 patients' motor states in daily-living conditions was collected without requesting specific motor tasks. 219 hours of wearable sensor data are collected using Microsoft Band 2 [1]. Acceleration and gyroscope data are collected at a frequency of approximately 62.5Hz; acceleration data only are used for the PD motor state classification in this research, similar to previous works (e.g. [25], [10], [11], [7]). Since the sampling interval is often irregular due to wireless communication between the wearable sensor and the recording device, a mobile phone in this case, the data are resampled to a constant interval at 120Hz. As the output, one minute windows should have 7200 samples, however, some windows have fewer samples due to missing data. To generate same-length inputs, we crop the windows to use 58-second data instead of 60-second, that is, all inputs have 6960 samples rather than 7200.

The motor states of the PD patients are labeled for each one minute interval by a clinical expert during the collection of the wearable sensor data. From the 219 hours of data, 154 hours of data are labelled under direct observation of the expert. 50 hours of data collected during walking and laying activities are removed from the training data due to limited observation during these activities. Finally, *no-symptom* data are removed to simplify the problem into a two class classification problem of bradykinesia and dyskinesia motor states, because the goal of the research is to characterize data augmentation methods. As a result, 77 hours of two class data are used for evaluating various data augmentation methods with CNNs for this research. Note that two patients who only have a small amount of *no-symptom* data are removed during the above preprocessing, thus, experiments are conducted with the data of the remaining 25 patients.

The amount of collected data from each patient varies, from 9 minutes to 273 minutes. Also, 12 patients exhibit both motor states while 13 patients only exhibit either bradykinesia or dyskinesia. For a fair evaluation, the 5 patients who have relatively large samples of both states are selected as test patients while the other 20 patients are selected as training patients. In the end, 3530 minutes of data (bradykinesia:1715, dyskinesia:1815) from the 20 training patients served as the training data and 1090 minutes of data (bradykinesia:442, dyskinesia:648) from the remaining 5 patients served as the test data. Note that no other preprocessing, e.g., data normalization or smoothing, is applied because they may confound the data label and subsequent results (Table 1).

4.2 The CNN architecture

7-layer CNNs are used for evaluating the performance of the PD motor state classification with various data augmentation methods

Table 1: The PD patient data used for this research

	Train (20 subj)			Test (5 subj)		
	Brady	Dys	Total	Brady	Dys	Total
#. Data (min)	1715	1815	3530	442	648	1090

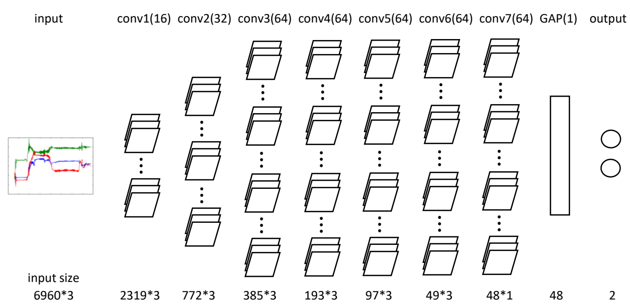

Figure 3: The CNN architecture used for the experiments. It consists of 7 convolutional layers with a global average pooling (GAP) layer at the end. For the cropping and sampling experiments, 2319*3 inputs are directly fed into the first layer rather than into the input layer.

(Figure 3). A reason for employing a CNN rather than a Long Short-term Memory (LSTM) [13], which is often used for learning time-series data, is to reduce the number of parameters to work with the small-scale PD data. In [28], a CNN has been successfully applied to wearable sensor data under the conditions of small temporal variability and large data availability. Thus, we may expect good performance with a CNN for PD motor state classification if the temporal localization variability of the data can be reduced and the amount of available data can be enlarged using appropriate data augmentation methods.

Each layer of the 7-layer CNN consists of 16-32-64-64-64-64-64 feature maps as presented in Figure 3. With strided convolutions, the sizes of the feature maps are changed to 2319*3, 772*3, 385*3, 193*3, 97*3, 49*3, 48*1 over layers, while the sizes of convolution filters of each layer are 4*1, 4*1, 3*1, 3*3, 3*3, 3*3, 3*1, respectively. Note that XYZ signals of the accelerometer are convolved in layers 4,5,6 and 7 to capture inter-vector-component features.

Fully connected layers which are frequently-used in CNN architectures for classification are substituted with a global averaging pooling (GAP) layer [21] to reduce the number of parameters. It is worth to reemphasize that learning from a small dataset with a large-parameter learning model tends to overfit, thus, it is important to employ a small-parameter model. By removing fully connected layers, we use a deep and sparse CNN model which is advantageous for preventing overfitting.

A convolutional layer, a batch normalization layer [14], and an activation layer form a single convolutional layer of the 7-layer CNN. Rectified units (ReLUs) are used for all activation layers except for the output layer which uses softmax units. For cropping and sampling data augmentation, 2319-sample windows are extracted from one minute windows and fed directly into the first convolutional layer, rather than into the input layer. That is, the CNN architectures for the cropping and sampling data augmentation experiments are 6-layer CNNs, which consist of 32-64-64-64-64-64 feature maps.

Table 2: The results of PD motor state classification with various data augmentation methods

(400 epoch)	Train	Test			
	Top	Top Aug	Top	Med	Std
No Aug	100.0	76.7	76.7	74.8	0.023
Jitter (0.01)	100.0	76.7	76.5	75.7	0.024
Jitter (0.05)	100.0	73.3	73.3	72.1	0.024
Jitter (0.1)	100.0	65.8	67.3	64.7	0.038
Scale (0.05)	100.0	81.2	80.9	78.9	0.024
Scale (0.1)	100.0	81.2	81.2	79.4	0.034
Scale (0.2)	100.0	81.4	81.2	80.2	0.029
Crop	97.0	79.1	71.9	68.9	0.023
Sample	99.2	75.6	72.6	72.4	0.026
Crop+Sample	98.6	81.9	76.8	70.3	0.070
Rotation	99.9	**88.6**	**88.6**	**87.3**	0.023
Permutation	99.3	79.5	78.7	74.5	0.023
Rot+Perm	97.0	**92.0**	**91.2**	**90.6**	0.039
Rt+Pm+Sc	95.7	**92.0**	**91.1**	**90.6**	0.032
Rt+Pm+Cr+Sp	92.7	89.4	88.3	86.8	0.069

4.3 Results

With the 3530 minutes of training data from 20 patients and the 1090 minutes of test data from 5 patients (Table 1), classification of PD motor states is performed using the CNN described in Section 4.2 (Figure 3) with the combination of various data augmentation methods proposed in Section 3.2 (Figure 2). All experiments are performed for 400 epochs and different training sets transformed with random hyperparameter values are used for each training epoch. For example, 3530 random rotation matrices are used for transforming the original training data for an epoch and a different 3530 random rotation matrices are used for another epoch.

The main results are presented in Table 2. Since the classification results can fluctuate based on the quality of the random augmentations of each epoch, we present the *Top* and *Median* results with standard deviations, *STD*, from the last 10 epoch results. Also, in addition to the data augmentation in the training dataset, the same data augmentation is performed on the test dataset to generate ×40 predictions; a single prediction, *Top Aug*, is then made by majority vote.

PD motor state classification with the original training data without data augmentation, *No Aug*, shows poor performance with an accuracy of 76.7%. This poor result is expected because of the challenges of the limited PD data as described in Section 3.1. *Jittering* with three different STD values (0.01, 0.05, 0.1) shows similar or worse performance than *No Aug*. Since jittering introduces rapid fluctuations which look similar to dyskinesia or tremor, the transformation does not preserve the labels, thus, gives a poor performance.

Interestingly, *scaling* the training data by multiplying with a random number sampled from normal distributions of zero mean and 0.05, 0.1, 0.2 STDs improves the classification performance to 81.2%, 81.2%, 81.4% in corresponding STD value orders. This result may imply that the absolute magnitudes of the data may be less important than the dynamic patterns of the data. In other words, it may indicate that bradykinesia and dyskinesia motor states are more easily distinguishable with the activity patterns of the data than the magnitude of the data.

Cropping and *sampling* data augmentation methods fail to improve the performance based on *Top* and *Med* results. A possible reason could be that they randomly drop the information of $\frac{2}{3}$ of samples, which could be a critical loss given the small dataset. The results of *cropping* and *sampling* also indicate that reducing the size of the inputs using strided convolutions is more effective than random cropping or sampling.

The best performance among the single data augmentation methods is achieved by the *rotation* method. This result indicates that a major source of variability is the pose of the wearable sensor. Arbitrary sensor placements and various hand poses may introduce a constant orientational offset and a large variability of the data. The result shows that this source of variability can be effectively compensated by the rotational data augmentation method. The performance improvement by rotational data augmentation may be particularly salient in the PD motor state classification problem because the states can be evaluated independently of the patient's hand pose.

The best performance among all experiments is achieved by the combination of *rotation* and *permutation* methods. This combination achieves 92.0% accuracy, which is 15.3 percentage points better than the *No Aug* result. The additional improvement from the result of *Rot* is obtained by diminishing the temporal variability using the permutational data augmentation method as described in Section 3.2. Another reason for the good performance could be a regularization effect by the complex data augmentation. Considering the training accuracy curves in Figure 4(a), the training curves of no augmentation and simpler augmentation cases are more easily saturated than the *Rot+Perm* curve and the test curve of *Rot+Perm* in (b) shows a better generalization performance with a regularization effect by the augmentation.

Rotations and permutations of wearable sensor data can be considered as label-preserving transformations in orientational space and in time space, respectively. Analogous to an object in an image that can be located in any XY position, a motor state pattern may be located in any temporal and orientational position, thus, rotation and permutation could be label-preserving transformations. Applying a constant rotation is not likely to change the labels because motions are likely to appear as time-varying patterns rather than being governed by constant offsets. This could be the reason for the different results between *scaling* and *jittering* because *scaling* applies a constant transformation to all samples, thus, it does not change the label, while jittering applies random variations to each sample, thus, it may change the label.

In short, data augmentation with *rotation* and *permutation* brings the best performance improvement while *rotation* is the most influential single data augmentation method for wearable sensor-based PD monitoring. *Jittering* is not effective because it may change the labels while *cropping* and *sampling* may cause too much information loss. *Scaling* improves the performance because it introduces constant transformations to all samples, but its effectiveness may vary in other tasks.

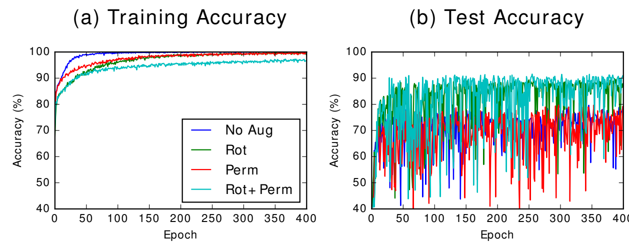

Figure 4: Training curves for *No Aug, Rot, Perm,* and *Rot+Perm.* The curve of *Rot+Perm* show that complicated data augmentation makes training more difficult while gives better test performance by a regularization effect.

All 87 incorrect test predictions and 200 randomly selected correct predictions out of 1003 correct predictions with the *Rot+Perm* augmentation method are presented in Figure 5. The data windows with a white background are bradykinesia-labelled data while the data windows with a yellow background are dyskinesia-labelled data. It can be seen that many examples of wrong predictions may be caused by disagreement-labels while very challenging PD data are correctly predicted using the 7-layer CNN with the *Rot+Perm* data augmentation.

5 CONCLUSION

In this paper, an automatic classification algorithm for PD motor state monitoring is developed based on wearable sensor data. PD motor state classification is a challenging task because of limited data availability, large inter-patient and inter-class variability, noisy labels, and interference by irrelevant motion signals. The challenging task is successfully tackled using a 7-layer CNN and the combination of rotational and permutational data augmentation methods, improving the baseline performance of 76.7% accuracy to 92%. Systematic experiments with various data augmentation methods provide a direction towards a general approach for augmentation for wearable sensor data.

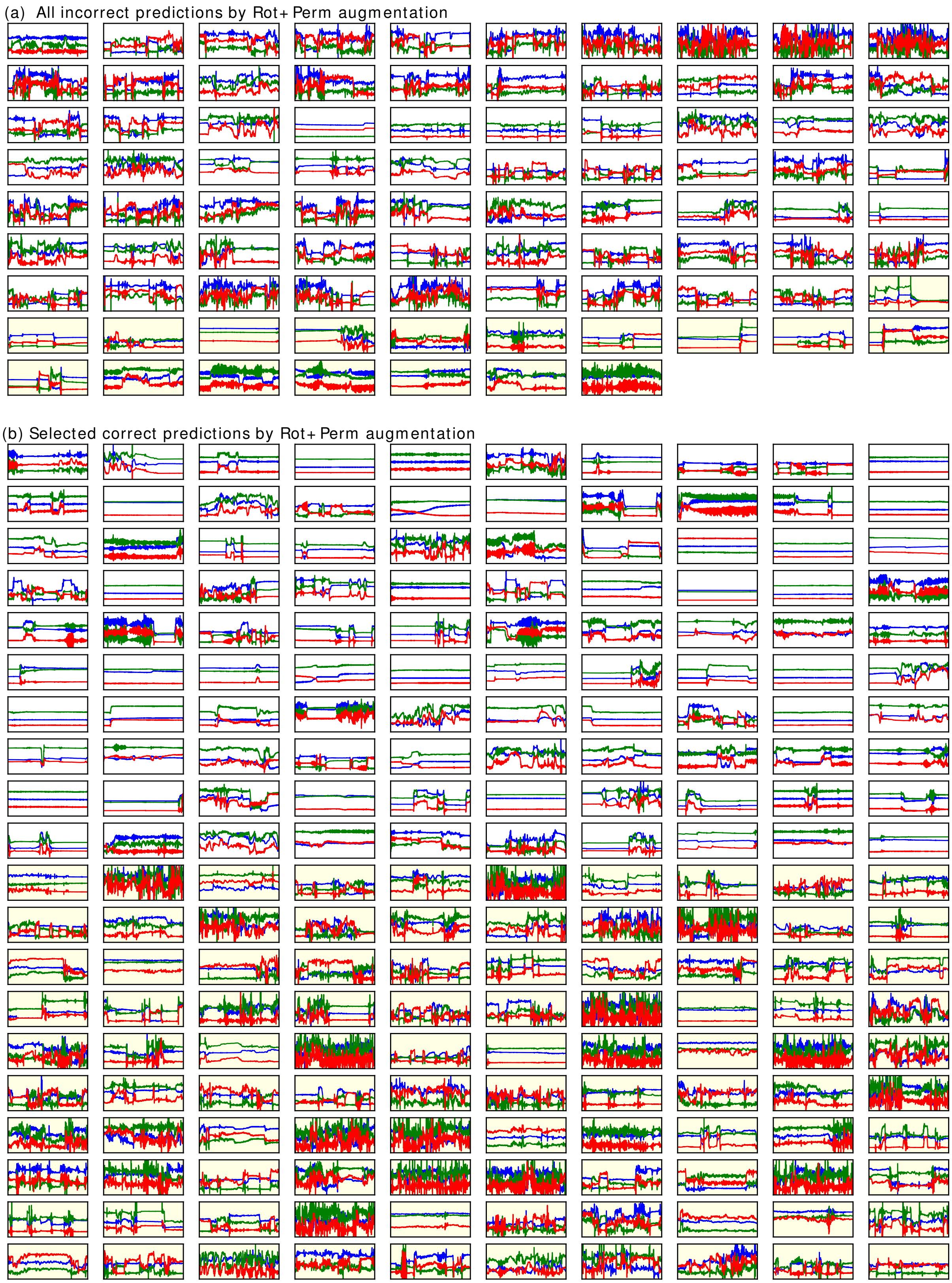

Figure 5: (a) 87 incorrect test predictions and (b) 200 randomly selected correct test predictions out 1003 correct predictions. White and yellow background color mean bradykinesia and dyskinesia true labels, respectively.

REFERENCES

[1] Microsoft Band 2. 2015. (2015). https://www.microsoft.com/microsoft-band/

[2] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. 2014. Return of the Devil in the Details: Delving Deep into Convolutional Nets. *CoRR* abs/1405.3531 (2014). http://arxiv.org/abs/1405.3531

[3] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning* 20, 3 (1995), 273–297. https://doi.org/10.1007/BF00994018

[4] Xiaodong Cui, Vaibhava Goel, and Brian Kingsbury. 2015. Data Augmentation for Deep Neural Network Acoustic Modeling. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.* 23, 9 (Sept. 2015), 1469–1477. https://doi.org/10.1109/TASLP.2015.2438544

[5] George E. Dahl, Dong Yu, Li Deng, and Alex Acero. 2012. Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 1 (Jan 2012), 30–42. https://doi.org/10.1109/TASL.2011.2134090

[6] Silvia Del Din, Alan Godfrey, Claudia Mazz, Sue Lord, and Lynn Rochester. 2016. Free-living monitoring of Parkinson's disease: Lessons from the field. *Movement Disorders* 31, 9 (2016), 1293–1313. https://doi.org/10.1002/mds.26718

[7] Bjoern M. Eskofier, Sunghoon I. Lee, Jean-Francois Daneault, Fatemeh N. Golabchi, Gabriela Ferreira-Carvalho, Gloria Vergara-Diaz, Stefano Sapienza, Gianluca Costante, Jochen Klucken, Thomas Kautz, and Paolo Bonato. 2016. Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson's disease assessment. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 655–658. https://doi.org/10.1109/EMBC.2016.7590787

[8] Ross Girshick. 2015. Fast R-CNN. In *The IEEE International Conference on Computer Vision (ICCV)*.

[9] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

[10] Robert I. Griffiths, Katya Kotschet, Sian Arfon, Zheng Ming Xu, William Johnson, John Drago, Andrew Evans, Peter Kempster, Sanjay Raghav, and Malcolm K Horne. 2012. Automated assessment of bradykinesia and dyskinesia in Parkinson's disease. *Journal of Parkinson's disease* 2, 1 (2012), 47–55.

[11] Nils Y. Hammerla, James M. Fisher, Peter Andras, Lynn Rochester, Richard Walker, and Thomas Plotz. 2015. PD Disease State Assessment in Naturalistic Environments Using Deep Learning. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15)*. AAAI Press, 1742–1748. http://dl.acm.org/citation.cfm?id=2886521.2886562

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. *CoRR* abs/1512.03385 (2015). http://arxiv.org/abs/1512.03385

[13] Sepp Hochreiter and Jörgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (Nov 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[14] Sergey Ioffe and Christian Szegedy. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ArXiv e-prints* (Feb. 2015). arXiv:cs.LG/1502.03167

[15] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. 2014. Large-scale Video Classification with Convolutional Neural Networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[16] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014*. Doha, Qatar, 1746–1751.

[17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 1097–1105.

[18] Ken J. Kubota, Jason A. Chen, and Max A. Little. 2016. Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures. *Movement Disorders* 31, 9 (2016), 1314–1326. https://doi.org/10.1002/mds.26693

[19] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (28 May 2015), 436–444. http://dx.doi.org/10.1038/nature14539 Insight.

[20] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng. 2009. Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML '09)*. ACM, New York, NY, USA, 609–616. https://doi.org/10.1145/1553374.1553453

[21] Min Lin, Qiang Chen, and Shuicheng Yan. 2013. Network In Network. *ArXiv e-prints* (Dec. 2013). arXiv:1312.4400

[22] João Massano and Kailash P. Bhatia. 2012. Clinical Approach to Parkinson's Disease: Features, Diagnosis, and Principles of Management. *Cold Spring Harb Perspect Med* 2, 6 (Jun 2012), a008870. https://doi.org/10.1101/cshperspect.a008870 a008870[PII].

[23] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. 2015. Is Object Localization for Free? - Weakly-Supervised Learning With Convolutional Neural Networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[24] Christiana Ossig, Angelo Antonini, Carsten Buhmann, Joseph Classen, Ilona Csoti, Björn Falkenburger, Michael Schwarz, Jürgen Winkler, and Alexander Storch. 2016. Wearable sensor-based objective assessment of motor symptoms in Parkinson's disease. *Journal of Neural Transmission* 123, 1 (2016), 57–64. https://doi.org/10.1007/s00702-015-1439-8

[25] Shyamal Patel, Konrad Lorincz, Richard Hughes, Nancy Huggins, John Growdon, David Standaert, Metin Akay, Jennifer Dy, Matt Welsh, and Paolo Bonato. 2009. Monitoring Motor Fluctuations in Patients With Parkinson's Disease Using Wearable Sensors. *IEEE Transactions on Information Technology in Biomedicine* 13, 6 (Nov 2009), 864–873. https://doi.org/10.1109/TITB.2009.2033471

[26] Tara N. Sainath, Abdel rahman Mohamed, Brian Kingsbury, and Bhuvana Ramabhadran. 2013. Deep convolutional neural networks for LVCSR. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. 8614–8618. https://doi.org/10.1109/ICASSP.2013.6639347

[27] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15 (2014), 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html

[28] Terry Taewoong Um, Vahid Babakeshizadeh, and Dana Kulić. 2016. Exercise Motion Classification from Large-Scale Wearable Sensor Data Using Convolutional Neural Networks. *CoRR* abs/1610.07031 (2016). http://arxiv.org/abs/1610.07031

[29] Pascal Vincent Roland Memisevic Xavier Bouthillier, Kishore Konda. 2015. Dropout as data augmentation. *ArXiv e-prints* (June 2015). arXiv:stat.ML/1506.08700