Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200289

The Smart Device System for Movement Disorders: Preliminary Evaluation of Diagnostic Accuracy in a Prospective Study

Julian VARGHESE [a,1], Michael FUJARSKI [a], Tim HAHN [b], Martin DUGAS[a], Tobias WARNECKE [c]
[a] *Institute of Medical Informatics, University of Münster, Germany*
[b] *Department of Psychiatry, University Hospital of Münster, Germany*
[c] *Department of Neurology, University Hospital of Münster, Germany*

Abstract. Consumer wearables can provide objective monitoring of movement disorders and may identify new phenotypical biomarkers. We present a novel smartwatch-based prototype, which is implemented as a prospective study in neurology. A full-stack Machine Learning pipeline utilizing Artificial Neural Networks (ANN), Random Forests and Support Vector Machines (SVM) was established and optimized to train for two clinically relevant classification tasks: First, to distinguish neurodegenerative movement disorders such as Parkinson's Disease (PD) or Essential Tremor from healthy subjects. Second, to distinguish specifically PD from other movement disorders or healthy subjects. The system was trained by 318 samples, including 192 PD, 75 other movement disorders and 51 healthy participants. All models were trained and tested with hyperparameter optimization and nested cross-validation. Regarding the more general first task, the ANN succeeded best with precision of 0.94 (SD 0.03) and recall of 0.92 (SD 0.04). Concerning the more specific second task, the SVM performed best with precision of 0.81 (SD 0.01) and recall of 0.89 (SD 0.04). These preliminary results are promising as compared to the literature-reported diagnostic accuracy of neurologists. In addition, a new data foundation with highly structured and clinically annotated acceleration data was established, which enables future biomarker analyses utilizing consumer devices in movement disorders.

Keywords. Mobile applications, artificial intelligence, machine learning, movement disorders

1. Introduction

Movement Disorders as Parkinson's Disease (PD) are primarily diagnosed via clinical examination. A meta-analysis by Rizzo et al. 2016 indicated that the pooled diagnostic accuracy is at 73.8% for general neurologists, geriatricians, or general practitioners (95% Credible Interval (CRI): 67.8%-79.6%) and 79.6% (95% CrI 46%-95.1%) for movement disorder experts regarding initial assessment and 83.9% (95% CrI 69.7%-92.6%)

[1] Corresponding Author, Julian Varghese, Institute of Medical Informatics, University of Münster; E-mail: julian.varghese@uni-muenster.de

regarding refined diagnosis after follow-up [1]. It ascertains that the overall validity of clinical diagnosis did not improve in the last 25 years and new biomarkers are needed, ideally with high availability. We present a novel mobile prototype, called the smart device system (SDS), which has been implemented and approved for a two-year prospective study [2]. The system is the first one, which captures two-side synchronous acceleration data from smartwatches with the integration of electronic questionnaires to account for early symptoms, medication, medical and family history. Health informaticians and movement disorder specialists designed the system, including a smartphone app to guide the examination step by step. Previously, we have demonstrated the accuracy of raw measurement data and tremor frequency detection as proof of concept to visualize and detect subtle movement characteristics, which are imperceptible to human vision [3].

As a clinical use case, we trained a diagnosis classification model. To our knowledge, most of the related models (summarized in [4]) for PD deal with limited sample sizes ($n<<100$) and are overly restricted to classify PD vs healthy. However, in clinical practice, the neurologist does not only face people being exclusively a PD case or healthy case. Instead, there are many differential diagnoses with PD-similar symptoms. In these cases, the models mentioned above would not be applicable or would perform significantly poorer. To address this issue, an important disease category is part of the patient recruitment: movement disorders other than PD as essential tremor, atypical parkinsonism, secondary causes of parkinsonism, dystonia, which will be summarized as DD in the following. Our system is trained for two clinically relevant classification tasks. First, to distinguish between movement disorders (PD+DD) and healthy subjects. Such a system could potentially be used at home-based settings or at general practices, e.g. to indicate whether certain abnormal movement characteristics (e.g. hand tremor) is pathologic or still normal (e.g. physiological tremor). Second, to distinguish PD from the rest (DD + healthy), which addresses a setting in specialized practices.

2. Methods

Details of the study design, examination steps and feature extraction have been published previously [2]. Study participant's assessment starts with smartphone-based data entry on medication, current diagnosis, family history and early symptoms. The smartwatch-based examination consists of ten different coordination tasks, including serials seven to provoke and monitor different movement characteristics while seated in an armchair. Altogether, the examination takes 15 minutes. The system is composed of two smartwatches (Apple Watch Series 4) for two-hand acceleration measurement and two smartphones (Apple iPhone version 7) for electronic questionnaires, examination guidance, watch-pairing and data transfer. These devices are connected via one central iOS app on the master smartphone. Sampling rate of acceleration data in all three axes is 100 Hz. Each sample is pseudonymized and ready for real-time visualization and further data processing [3]. Table 1 summarizes the three aforementioned relevant study participants categories PD, DD and Healthy. The full list with demographic details and years of disease onset is provided for each sample in the supplement. It also includes diagnosis details of the DD cluster [5]. All diagnoses were confirmed by neurologists and finally reviewed by one senior movement disorder expert. The current stack of the data analytics pipeline is implemented in Python, illustrated in Figure 1 and includes

Artificial Neural Networks (ANN), Random Forests and Support Vector Machines (SVM) as the three main classifiers.

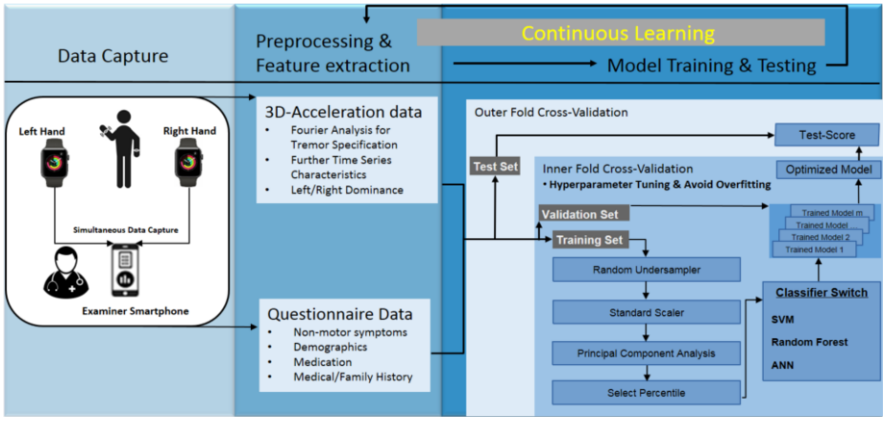

Figure 1. Data Analytics pipeline. ANN = Artificial Neural Networks, SVM= Support Vector Machine.

Table 1. Class population.

Class	Sample Size	Average Age (SD)
PD	192	65.71 (9.61)
DD	75	60.88 (15.58)
Healthy	51	60.45 (14.22)

TSFresh 0.12.0 is used to extract a series of general-purpose time-series characteristics [5]. Moreover, acceleration data is processed with Fast Fourier Transform to extract tremor frequencies. Amplitude-Histogram analyses are applied and checked for dominances on the left or right-hand side. A complete list of all extracted features and descriptions is given in the supplement [6]. The set of features is trained with nested cross-validation using five inner and five outer folds. During one inner fold, we applied the random undersampler from Scikit Learn 0.21.3 [7] in order to remove the bias towards the majority class by randomly removing samples of that set. The standard scaler from Scikit Learn subtracts the mean and scales to the unit variance for every feature. The principal component analysis (PCA) reduces the dimensionality without assuming independence. The Scikit Learn-based 'Select Percentile' step randomly selects a subset of features, which are then used for training the classifier. While the standard Scikit Learn library implements SVM and Random Forests, the ANN is implemented with dense multi-layered perceptron networks using the KERAS 2.2.4 package and Google's Tensorflow 1.3.1, which provides full GPU support [8]. The whole process of model training and testing is wrapped by Photon-AI, a hyperparameter optimization toolbox [9]. During an inner fold, we optimize the hyperparameters for the PCA, the Select Percentile and the classifiers. We used the Scikit Learn's grid-search optimizer as our optimizer. A list of the full hyperparameter space is provided in our supplement [6].

3. Results

Tables 2 and 3 list the best performing models for each of the three classifiers after nested cross-validation. While the ANN performed best for the first task regarding average accuracy of 0.89 and precision of 0.94, the SVM – using radial basis function (RBF) as kernel – performed best in the second task with accuracy of 0.79 and precision of 0.81. None of the different classifiers outperformed the others when taking the overlap of their corresponding standard deviations into account. Details on all optimized hyperparameter values and best performing features are provided in the supplement [6]. Further classifiers, including deep neural networks, are being evaluated and discussed in the next section.

Table 2. Average Performances for general classification task: Parkinson's and related movement disorders vs Healthy. SD= standard deviation. Rbf= Radial Basis function.

Estimator	Accuracy (SD)	Precision (SD)	Recall (SD)	F1 (SD)
ANN	**0.89** (0.00)	0.94 (0.03)	0.92 (0.04)	0.93 (0.00)
Random Forest	0.88 (0.04)	0.91 (0.04)	0.95 (0.04)	0.93 (0.02)
SVM - rbf	0.89 (0.03)	0.92 (0.03)	0.95 (0.04)	0.93 (0.02)

Table 3. Average Performances for specific classification task: PD vs. Other related movement disorders and Healthy.

Estimator	Accuracy (SD)	Precision (SD)	Recall (SD)	F1 (SD)
ANN	0.77 (0.04)	0.78 (0.04)	0.90 (0.06)	0.83 (0.03)
Random Forest	0.75 (0.03)	0.75 (0.03)	0.91 (0.04)	0.823 (0.02)
SVM - rbf	**0.79** (0.01)	0.81 (0.01)	0.89 (0.04)	0.85 (0.01)

4. Discussion

The more general task was evaluated with precision and recall above 90% for all classifiers, which highlights potential applicability in non-specialized settings as home-based assessment or general practices. Regarding the more specific task, the average diagnostic accuracy (79%) of the best performing model is higher than pooled reported accuracy of general practitioners and comparable to movement disorder experts regarding initial assessment [1]. A number of machine-learning approaches share a major disadvantage as they are trained by retrospective data sources, with high data quality, which however might be illusory at real-time or point of care. A major strength of our mobile system is that it only requires the self-generated data from the 15 minutes examination being not dependent on any further patient data or external information systems. Thus, we are confident that our results are reproducible in prospective scenarios. The reported results still need to be taken with caution as our setting comes with general limitations: The study is conducted at a single site, and thus we cannot rule out bias due to specific site population, examination setting or target diagnoses. We recruit from a large tertiary care centre and include all types and stages of Parkinson's disease and other movement disorders, and therefore, we believe the training data provides a suitable data foundation to start from. However, as a matter of our recruitment

setting, only a minority of samples are Parkinson's cases with early disease stage, which would be the ideal population to train from. After study completion, a multi-centric study design is planned for 2021, which would include different outpatient clinics and home-based assessments. At the current state, the system is to be used for scientific purpose only as it is not cleared as a medical device.

We have tested several approaches towards Deep Learning with recurrent networks as Long-Short-Term Memories. Their performances were remarkably poorer, being not significantly higher than random guessing. Our sample size may not be sufficient (e.g. > 10,000) to train a deep network that requires substantially more model parameters. As our data-preprocessing steps evolve and the sample size grows, we are monitoring different types of Neural Networks and further traditional statistical models. Apart from the machine-learning framework, the highly interactive research platform enables researchers to visualize objective movement characteristics and to have a deeper look at subtle characteristics within a hundredth of a second. In a previous study, we could visualize tremor characteristics imperceptible to non-sensor-based human vision [3]. This could unveil further morphological features in the future, which are not yet utilized.

To conclude, we have generated a sensor-based data foundation to study movement characteristics in a broad area of different movement disorders. Machine Learning classifiers show high accuracy for distinguishing relevant diagnoses classes compared to expert-based performances. Further research is underway to show generalizability of the system.

5. Acknowledgement

This work is funded by the Innovative Medical Research Fund (Innovative Medizinische Forschung, I-VA111809) of the University of Münster.

References

[1] G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, G. Logroscino, Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis, *Neurology* **86(6) (**2016), 566–76.

[2] J. Varghese, S. Niewöhner, I. Soto-Rey, S. Schipmann-Miletić, N. Warneke, et al., A Smart Device System to Identify New Phenotypical Characteristics in Movement Disorders, *Front Neurol* **10** (2019), 48.

[3] J. Varghese, S. Niewöhner, M. Fujarski, I. Soto-Rey, A.L. Schwake, et al., Smartwatch-based Examination of Movement Disorders: Early Implementation and Measurement Accuracy, *EGMS* (2019). DOI:10.3205/19GMDS136.

[4] AU. Haq, J.P. Li, M.H. Memon, J. Khan, A. Malik, et al., Feature Selection Based on L1-Norm Support Vector Machine and Effective Recognition System for Parkinson's Disease Using Voice Recordings, *IEEE Access* **7** (2019), 37718–34.

[5] M. Christ, N. Braun, J. Neuffer, A.W. Kempa-Liehr, Time series feature extraction on basis of scalable hypothesis tests (tsfresh - a python package), *Neurocomputing* **307** (2018), 72-7.

[6] J. Varghese, *Online Supplement SDS MIE 2020,* https://uni-muenster.sciebo.de/s/dYv1d7weDar7xnZ, (accessed 09/31/2019).

[7] G. Hackeling, *Mastering Machine Learning with scikit-learn*, Packt Publishing, 2017.

[8] *Keras Documentation,* https://keras.io, (accessed 10/01/2019).

[9] *Photon-AI. A Python-based Hyperparameter Optimization Toolbox,* https://www.photon-ai.com/, (accessed 30/09/2019).