Defining Optimal Exercises for Efficient Detection of Parkinson's Disease Using Machine Learning and Wearable Sensors

Aleksandr Talitckii, Anna Anikina, Ekaterina Kovalenko, Aleksei Shcherbak, Oscar Mayora, Olga Zimniakova, Ekaterina Bril, Maxim Semenov, Dmitry V. Dylov, *Member, IEEE*, and Andrey Somov

***Abstract*— Our society exhibits a worldwide trait of a quickly growing cohort of patients with neurodegenerative diseases, such as Parkinson's disease (PD). According to the analysts, there is a plausible "PD pandemic" to occur within the next two decades. Nowadays, the research in the area focuses on how to detect, predict, or classify PD and similar diseases without addressing the point of what activities or exercises a subject should do to improve the performance of these tasks. In this article, we propose a method based on machine learning (ML) and wearable sensors to identify the optimal exercises for the efficient detection of PD in patients. We first define 15 common tasks that are typically used to diagnose PD in modern clinical practice. However, these exercises still carry a high risk of misdiagnosis and, moreover, not all of them work well in the scope of existing ML solutions to support the diagnosis. Herein, we collect the data in a real clinical setting using a compact wearable wireless sensor node entailing a board gyroscope, accelerometer, and magnetometer. Application of ML methods to the collected data reveals three "most efficient" exercises to assist diagnosticians with the highest discriminating power (0.9 ROC AUC in each task). The proposed solution can be implemented as a medical decision support system for real-time PD diagnostics.**

***Index Terms*— Machine learning (ML), Parkinson's disease (PD), wearable sensing.**

I. Introduction

PARKINSON'S disease (PD) is a progressive neurological disease caused by the destruction and the death of neurons. Worldwide, the number of cases of the disease varies from 41 per 100 000 people after 40 years old to 1903 cases per 100 000 people after 80 years old [1]. PD is more common in men (1729 cases per 100 000 people) than women (1644 per 100 000) [1]. The World Health Organization expects that by 2040 the number of patients with PD will surpass cancer as the second leading cause of death. Moreover, PD is not treatable, with the mere inhibition of the process being the only option for this disease. Therefore, it is vital to identify PD at the earliest stage and to get feedback about the therapy.

Diagnosis of PD is based on clinical features, however, many similar diseases could be confused with PD by their symptoms [2]. Oftentimes, the initial diagnosis is incorrect and only the observation of the patient over time helps a doctor to find the correct diagnosis. In some patients, the correct diagnosis is found only after a postmortem examination [3]. Moreover, about 25% of older people have symptoms of PD without the actual PD [4]. In practice, physicians use Movement Disorder Society Clinical Diagnostic Criteria for PD (MDS-PD). According to these guidelines, the diagnostic decision is primarily dependent on the presence of bradykinesia along with the rest tremor in the patient [4].

At the moment there are three widely used approaches for PD detection through sensing: 1) *Sleep analysis*—This approach relies on the recognition of typical signs of PD while a subject is asleep [5]. In particular, this method performs the analysis of the movement of the eyes in sleep. However, it is a time-consuming process requiring the complex installation of equipment above a sleeping person for detection of rapid-eye movements [6]. 2) *Speech analysis*—Change in speech ability is another sign of progressing PD [7]. In this case, machine learning (ML) is often used by signal analysis methods for figuring out the variations in speech. The sophistication of the process consists of filtering the collected data and eliminating the noise. In addition, scientists believe that this technique can be used to supplement the other approaches to increase the accuracy of the disease diagnosis. 3) *Behavior analysis*—A set of these approaches rely on the wearable sensors [8] or the video [9]. Although the video analysis demonstrates reasonable accuracy of PD detection [10], it still imposes strict requirements on the illumination conditions and the clearance within the field of view of the camera (local objects frequently interfere). In contrast to the video approach, wearable sensing is easier in terms of data collection. Moreover, the sensors can be used continuously and uninterruptedly, with the analysis of collected data being realized in the real-time scenario and not interfering with the regular activities of the subject. Most of the state-of-the-art methods use an accelerometer as

Manuscript received April 5, 2021; revised June 8, 2021; accepted July 6, 2021. Date of publication July 16, 2021; date of current version July 28, 2021. The Associate Editor coordinating the review process was Dr. Datong Liu. *(Aleksandr Talitckii, Anna Anikina, and Ekaterina Kovalenko contributed equally to this work.) (Corresponding author: Andrey Somov.)*

Aleksandr Talitckii, Anna Anikina, Ekaterina Kovalenko, Aleksei Shcherbak, Dmitry V. Dylov, and Andrey Somov are with the Center for Computational and Data-Intensive Science and Engineering (CDISE), Skolkovo Institute of Science and Technology, 121205 Moscow, Russia (e-mail: a.somov@skoltech.ru).

Oscar Mayora is with the eHealth Lab, Bruno Kessler Institute, 38122 Trento, Italy.

Olga Zimniakova, Ekaterina Bril, and Maxim Semenov are with A. I. Burnazyan Federal Medical and Biophysical Center, 123098 Moscow, Russia.

Digital Object Identifier 10.1109/TIM.2021.3097857

the only sensor, aiming at detecting the tremor at specific frequencies [11]. It may result in false PD detection because the tremor can be the sign of another disease (e.g., thyroid disorders).

Currently, there are numerous experiments reported in the literature which aspire to aid the detection of PD, with the focus ranging from sensing, positioning of the sensors, and the ultimate data analysis. However, there is still a lack of research for the exercises that enable efficient diagnostics of PD. Especially, considering the inexpensive technologies, such as the off-the-shelf sensors and the open-source ML methods. Indeed, the growing population of patients with neurological diseases, which is getting "younger" [12], stimulates the research community to consider these methods for prompt identification of the first symptoms.

In this article, we significantly extend our preliminary research results published as a short article [13] by defining the optimal exercises for efficient detection of the PD using ML and wearable sensors to improve the PD diagnostics. For this reason, we demonstrate the full framework for PD detection. In particular, in collaboration with the neurologists, we identify 15 exercises that are typically used to detect the signs of PD. Next, we perform the data collection in real settings with the subjects and data analysis to figure out which exercises guarantee the most accurate PD detection. To the best of our knowledge, this investigation was not performed before and there was no attempt to define the efficient exercises through the use of wearable sensors and ML. Apart from improving efficiency, the proposed approach helps to reduce the time taken to diagnose, ensuring efficient high-quality care, and moreover, to assist in reducing the hospitalization costs. Generally speaking, the approach proposed in the following is in the scope of healthcare Internet of Things [14], allowing the subjects to easily perform just a few exercises for providing continuous feedback to the neurologists and monitoring the status of subjects progress automatically.

The article is organized as follows. In Section II, we present comprehensive background on the PD and discuss relevant research works in the area. In Section III, we present the methodology used in this work. We demonstrate the data analysis in Section IV and obtained results in Section V. Finally, we provide concluding remarks and discuss our future work in Section VI.

II. Background

A. Parkinson's Disease

There are several symptoms that characterize the PD [15]:

1) Bradykinesia (i.e., slowness of movement) and difficulty in walking;
2) Rest tremor;
3) Rigidity;
4) Hypokinesia (i.e., poverty of movement);
5) Dementia in later stages.

Current symptomatic drugs provide significant relief of PD symptoms, but the use of such drugs is the most effective in the early stages of the disease until a significant loss of neurons occurs [16]. In practice, at least 50% of the nigrostriatal neurons are already lost at the time of diagnosis. In addition, significant difficulty in making a diagnosis is the existence of a number of neurological diseases, which in their symptoms are similar to PD. In addition, there is a special term for determining the group of conditions that have similar symptoms of PD—Parkinsonism. PD is a special case of Parkinsonism, there are also the following diseases: supranuclear palsy, multiple system atrophy, Lewy Body Disease or Dementia with Lewy Bodies, corticobasal degeneration, corticobasal ganglionic degeneration, and so on [17]. Approximately 15% of people with Parkinsonism have other diseases than PD. Thus, we conclude that PD's original criteria have their own exceptions, e.g., bradykinesia and tremor.

Bradykinesia is a slowness of movement and decreased movement amplitude. Bradykinesia is visible when moving the limbs (for example, tapping fingers). In addition, it is important to note that bradykinesia is not always caused by the PD, and the main feature of PD is bradykinesia that occurs during the continued movement [4].

Bradykinesia tremor has its own specific features that separate PD from other diseases. Only rest tremor is a sign of PD, while another type- action tremor (kinetic and postural tremors) is not relevant for parkinsonism. An important part is the correct diagnosis of rest tremor (as well as action tremor). The manifestation of rest tremor occurs when the patient is resting, and it also occurs when the patient maintains a certain position for a long time. Physicians cannot consider the only tremor since there are cases when the rest tremor is observed in patients with a different diagnosis than PD [18].

B. Related Work

1) Wearable Sensing: Sensing devices are often fixed on a subject body and are typically used for tremor and fall detection and gait instability analysis [19], [20]. The sensor nodes are low-power devices equipped with a microcontroller, the sensor(s), a wireless transceiver, and a battery. Since the microcontroller does not perform any complex mathematical operations the collected data are sent to the PC or server for further data analysis [21].

Gugliandolo *et al.* [22] have reported on a wearable sensing device for tremor detection able to collect, store, and transmit the data collected from the array of accelerometers via Wi-Fi. In addition, the proposed device is equipped with a global positioning system for tracking the subject location, which is obviously enabled in outdoor scenarios. We note here that Shawen *et al.* [23] demonstrated that one accelerometer and consequent data analysis relying on ML is enough for PD detection and various classification tasks. Another wearable system for tremor detection has been proposed by Heijmans *et al.* [24]. The system takes an advantage of two wearable sensors placed on a wrist and on a chest of tested subjects. These patients had to wear sensors during the daytime and charge the device during the nighttime. In addition to collecting the sensor data, the subjects also had to fill in the questionnaires from an application installed on the smartphone. Data were collected into the internal device's memory, time stamped related to questionnaires, and extracted

after the measurement period. The authors used a 200-Hz sample rate for the data collection and a 3-Hz low-pass filter.

A more advanced sensing solution is proposed by Pierleoni *et al.* [25] and report on a smart inertial system for 24-h monitoring of patient condition for classification of tremor and freezing gait in PD. The authors developed their own wearable sensor based on assembled separated modules and the sample rate of data collection was 512 Hz. Data from the accelerometer and gyroscope were sent via Bluetooth to a patient's smartphone with the installed application which then sent obtained data to the cloud where the data were processed by an algorithm helping detect the tremor and freezing of gait. Another sensing solution addressing the problem of postural stability in the scope of PD was proposed in [26]. The sensor node has a microcontroller, an accelerometer, and a wireless transceiver on board. The device is designed for wearing on the chest. However, the authors make a special focus on the postural instability classification and the device assessment in the reference motion analysis laboratory.

In addition, there are other wearable solutions for PD gait assessment, including the application of on-shoe wearable sensors [27] and usage of electromyography (EMG) sensors for gate freezing detection and muscle stimulation [28]. These solutions are personalized and more complicated in terms of sensors installation which may require medical assistance. In this work, we use an our-of-the-shelf wireless sensor node with customized software with the aim to simplify the data collection procedure.

2) Data Analysis: A wide range of research works in the area has been published in the literature primarily focusing on PD detection or PD prediction [29]. In this section, we discuss some of the relevant works.

Lahmiri [30] has statistically investigated the differentiation of walk of healthy young/elderly and PD patients. The research results have demonstrated the evidence of substantial group differences occurring in the measured gait nonlinear patterns. The author came to the conclusion that steadiness and persistence of walk were the most obvious factors of PD.

Barrantes *et al.* [31] proposed a smartphone-based system to differentiating PD and essential tremor. During experiments, the authors placed a smartphone over the hand dorsum and recorded data from the built-in smartphone's accelerometer 30 s at rest condition and the same duration at arm-stretching condition. Receiver operating characteristic (ROC) curves of total spectral power were used for detection of PD against healthy subjects and ROC curves of relative energy were used to detect PD against essential tremor. The authors consider their system for immediate results in an environment where more sophisticated diagnostic is unavailable.

Tahafchi and Judy [32] reported about using wearable EMG and inertial measurement unit sensors to estimate the number and severity of freezing of gait episodes of subjects with PD for determining the dose of levodopa and adjusting deep brain stimulation therapy. Sensors were placed on feet and shanks. EMG allows researchers to register muscle contraction when movement is absent and inertial measurement unit (IMU) sensors do not register anything. After data preprocessing, the authors applied a fully convolutional neural network (FCNN) with small network weights and dropout technique in the hidden layer to reduce overfitting because of the small size of the dataset for typical FCNN operation.

One more example of utilizing neural network belongs to Lonini *et al.* [8]. In this research, the authors did not see any substantial improvement of convolutional neural networks relative to the random forest.

Recent research by Shawen *et al.* [23] encompass the important questions of data collection methodology in PD and bradykinesia. As it turned out if only the accelerometer's data are used, the placement of sensors is not significantly important and the effectiveness of ML models is approximately the same whether the sensor is placed on the dorsum part of the hand or on the wrist. It is fair for both tremor detection and bradykinesia detection. However, the results of bradykinesia detection are more sensitive to gyroscope data in comparison with accelerometer data. Decreasing of data acquisition sample rate down to 5 Hz does not lead to significant impairment of bradykinesia detection results while decreasing sample rate below 30 Hz leads to the decreasing area under the ROC curve (AUC) curve of tremor detection. At sample-rate frequencies above approximately 30 Hz, the AUC curve of tremor detection tends to reach a plateau. These results are in accordance with the Nyquist sampling theorem and characteristic frequencies of tremors.

Locatelli and Alimonti [33] used a wearable sensor comprising an accelerometer, gyroscope, and magnetometer with an integrated quaternion-based sensor fusion algorithm for the orientation estimation. Data were collected with a 50-Hz sample rate from 27 subjects performing four special tasks. During the processing data were filtered by bandpass Chebyshev filter and then fast Fourier transform (FFT) was applied to each x, y, z, and scalar w components. The square root of the sum of the square of each FFT component was additionally calculated. Results show that frequency spectrum and amplitudes for essential tremor and for parkinsonian tremor are different. Besides this, the authors noted that each type of tremor in different tasks has a different degree of manifestation. From our point of view, it shows that different tasks may have different effects in detecting parkinsonian tremor and essential tremor.

Summarizing all, PD related to wearable sensors could be diagnosed with the help of a short test consisting of special exercises or due to monitoring of daily activity. There are reasons to suppose, that type of exercise matters in ML diagnostic. In this work, it was investigated which exercises are the best for detecting PD using wearable sensors and ML.

III. Methodology

In this work, we aim at the identification of exercises allowing for accurate PD detection.

For this reason, we solve the following problems.

1) *PD Stage Determination:* In this case, we defined only one and two stages of PD. It is important to understand what occurs in PD's early stages in terms of the main differences between these stages.

2) *Healthy and PD Subjects Classification:* We identify general features and characteristics of neurological disorders.
3) *Classification Between the PD and Other Related Neurological Disorders:* The difference between the PD and other neurological disorders is not significant. The application of ML techniques enables the detection of specific features belonging to the PD class only.

We designed a set of 15 exercises helping identify the symptoms specifically attributed to the PD. Exercises 1, 3, 4, 6, 7, and 12 were adopted from [8], while exercises 2, 5, 8, 9, 10, 11, 13, 14, and 15 came from the MDS-Unified PD Rating Scale. In addition, exercises 5, 8, 9, 13, 14, and 15 were specifically accentuated by the neurologists, as they were successfully used during the diagnostic procedures recently. Similar to [8] and following our recent research [13], the exercises used in this work are grouped into the four classes and are provided in Table I:

1) Gross Motor (GM)—Three Exercises #1, #3, and #6: Help assessing the subject movements which are often used in everyday activities.
2) Clinical Evaluation (CE)—Seven Exercises 2, 10, 11, 12, 13, 14, and 15): Include typical procedures used by the neurologists to reveal the tremor; these are specific movements that clinicians use to evaluate the additional PD symptoms.
3) Fine Motor (FM)—Four Exercises #4, #5, #7, and #9: Required for assessing the fine coordination of the subject.
4) Tremor at Rest—One Exercise #8: Assists in revealing the tremor, while the body is restrained.

Exercise 4 and exercise 5 are slight variations of the same task. Exercise 9 is technically three separate assignments; however, they were grouped into one as they are directly related to each other and simpler for the patient to do without a pause in between. While the subject was reading the sentence, audio was recorded for possible further analysis. Upon discussing the list of exercises with the neurologists, we came to the conclusion to limit this list and identify up to three exercises ensuring the best accuracy of PD detection. This is necessary to meet formal timing requirements per patient.

During our experiments, the SensorTile[1] IMU modules were used as the wireless wearable sensors for data collection, while the subject performs the common exercises (see Table I). Each of these compact modules is based on the STM32L476JG microcontroller unit and the BlueNRG-MS Bluetooth low energy (BLE) chip and supplied from the onboard Li-ion battery. There is a set of different sensors onboard each module including an accelerometer, gyroscope, and magnetometer. Each measurement was carried out with a frequency of 100 Hz and was sent via BLE to the Intel Next Unit of Computing (NUC) computer equipped with a special software coded using Python (see Fig. 1). For wearing the sensor node, we utilized a watchband to fix it on a hand where the watchband goes between the thumb and forefinger. In all the exercises, the patients wore the device attached to the hand, as shown in Fig. 1(b). This place is featured by the high amplitude of the tremor and allows for maximizing the signal-to-noise ratio of the sensor's data and the comfort of the subjects. Measurements from the accelerometer, gyroscope, and magnetometer were synchronized with each other, so we received a set of three time-stamped measurements in each sample. Data were collected separately for each exercise.

[1]http://www.st.com/

TABLE I

COMMON EXERCISES FOR SUBJECTS. (a) GM: GENERAL MOVEMENTS INVOLVING THE LARGER MUSCLES. (b) CE: SPECIFIC MOVEMENTS THAT CLINICIANS USE TO EVALUATE ADDITIONAL SYMPTOMS. (c) FM: MOVEMENTS INVOLVING SMALLER MUSCLES. (d) TR: TREMOR THAT OCCURS WHEN MUSCLES ARE RELAXED

Exercise	Purpose	Exercise #
Stride, take a seat and rise to standing position	BR	1
Towel folding	KT, BR	3
Move books from one organizer to another	KT, BR	6

(a)

Exercise	Purpose	Exercise #
Pronation/Supination of arms	KT	2
Tapping fingers on hard surface	KT, BR	10
Tapping index and thumb together while elbows are bent	KT, BR	11
Touch nose with index finger	AT, BR	12
Stand with arms stretched ahead	PT	13
Stand with arms folded in front of chest, palms pointed inward	PT	14
Stand with arms folded in front of chest, palms pointed outward	PT	15

(b)

Exercise	Purpose	Exercise #
Attach nut to bolt and tighten	KT, BR	4
Attach nut to anchor bolt and tighten	KT, BR	5
Pour water into glass	KT, BR	7
Pour water into glass	AT, BR	9

(c)

Exercise	Purpose	Exercise #
Relax arms	RT	8

(d)

Note: RT - Tremor at Rest, BR - Bradykinesia, KT - Kinetic Tremor, AT - Action Tremor, PT - Postural Tremor, .

In total, we collected the data from 83 subjects, 39 women, and 44 men, with the age ranging from 22 to 84 (52.9 $\pm$ 22.9) years. Out of these, 42 subjects had PD,

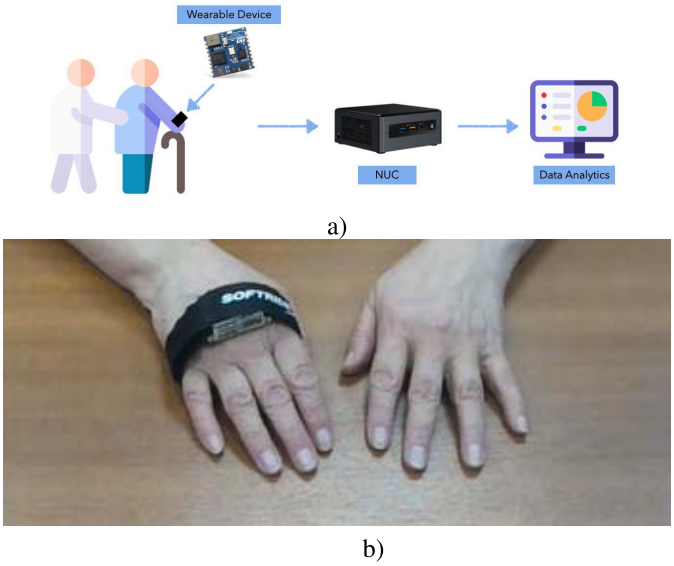

Fig. 1. Experimental test-bed for data collection: (a) measurement scenario and (b) wearable sensor.

17 subjects had other related neurological diseases presenting with tremors, and 24 subjects were healthy. For the PD patients, in particular, both the stages were presented in the following manner: stage 1 (8 patients), stage 2 (26 patients), stage 3 (7 patients), and stage 4 (1 patient). All patients performed each exercise to the best of their ability. In Fig. 2, we demonstrate the sensor data of healthy subjects and subjects with the PD.

IV. Sensor Data Analysis

A. Data Preprocessing

Data preprocessing is an important part of our work. First of all, we split the signal into segments of 5 s each so that the overlap will be 50%. After this operation, we got about 10 000 samples, which is 670 samples per exercise. The second stage of data processing was the application of a high-pass filter with 0.5-Hz frequency. This is necessary to remove the effect of a limb orientation. All parameters were selected according to [33]–[35].

B. Feature Extraction

The standard approach is to extract various features. First of all, we used statistics, such as the mean, standard deviation (STD), and other features (skew, kurtosis, minimum, and maximum values). For a more detailed analysis, we used the rolling window. We denoted the filtered data as the main trend and the remaining (signal without trend) we defined as the noise signal. For each sample, we found its spectrum using FFT. This approach is considered effective for the detection of PD and tremor [33].

Since parkinsonian tremor and bradykinesia have their own characteristics in frequency [33], the FFT of the signal was split into two classes: bradykinesia and tremor. For extracting the *tremor* related features, we used the frequency range from 3 to 10 Hz, while the *bradykinesia* related features were extracted in the frequency range from 0 to 3 Hz. We note here that the localization of bradykinesia-related signs is limited by 3 Hz [35].

We were looking for peaks in the frequency domain while considering the trend and noise signal. For each symptom (tremor and bradykinesia), we first calculated the mean and STD of all the amplitudes and frequencies of the full range of the spectrum (mean and STD of amplitudes) and then considered only the peaks where the amplitude was greater than the mean plus the STD of the Fourier transformed data. An example of FFT peak separation for bradykinesia and tremor is shown in Fig. 3. To extract useful features from the peaks, we calculated the mean, STD, min, and max values for peak frequencies and amplitudes and also found the dominant frequency. A general description of the features is provided in Table II. In total, we received about 500 different features describing diagnostics

$$N_f = \overset{N_s}{9} \times \overset{R}{2} \times \{\overset{SF}{8} + \overset{F}{2} \times \overset{\text{Peaks}}{8} + \overset{E}{2}\} = 468 \tag{1}$$

where N_f is the total number of features, N_s is the number of sensors (three axis and three sensors), SF are the standard statistical features, R is the trend or noise part of data, F is the frequency or amplitudes, $Peaks$ is the number of Fourier features, and E are the features of energy. Here, we calculated μ, σ λ_-, λ_+ for all frequencies, all frequency peaks, all amplitudes, and for all amplitude peaks. The power was calculated only for the peaks and for the entire spectrum.

For demonstrating the significance of various features, we used the Gini importance method (or mean decrease in impurity) [36]. It is often used for tree-based methods. According to the mean decrease in impurity, the best nine features are shown in Fig. 4. Names of features are intuitive. The first word presents the sensor data and the direction (x, y, or z) for which this feature was calculated. The second indicator is Brady or Tremor. It tells us in which range of frequency values was extracted. Finally, we wrote the part of a signal (trend or noise), and denoted amplitude or frequency and the statistic value that we extracted from data (minimum, maximum, or mean value). One can see that almost all the features are from the noise part of the data. This shows that the main trend features are redundant for the analysis of PD. Indeed, the main trend reflects the type of exercise and not the category of the patient.

C. Dimensionality Reduction and Features Selection

For dimensionality reduction (DR) problem, we applied the next algorithms.

1) Principal component analysis (PCA) featured by the sigmoid and polynomial kernel, linear, and radial basis function [37], [38].
2) Factor analysis (FA) [39].
3) Independent component analysis (ICA) [40].
4) Multidimensional scaling (MDS) [41].

Next, we refer to the number of components as the dimension of the new feature space similar to the PCA analysis. It demonstrated that the DR was characterized by the six best components representing over 90% of the total variance.

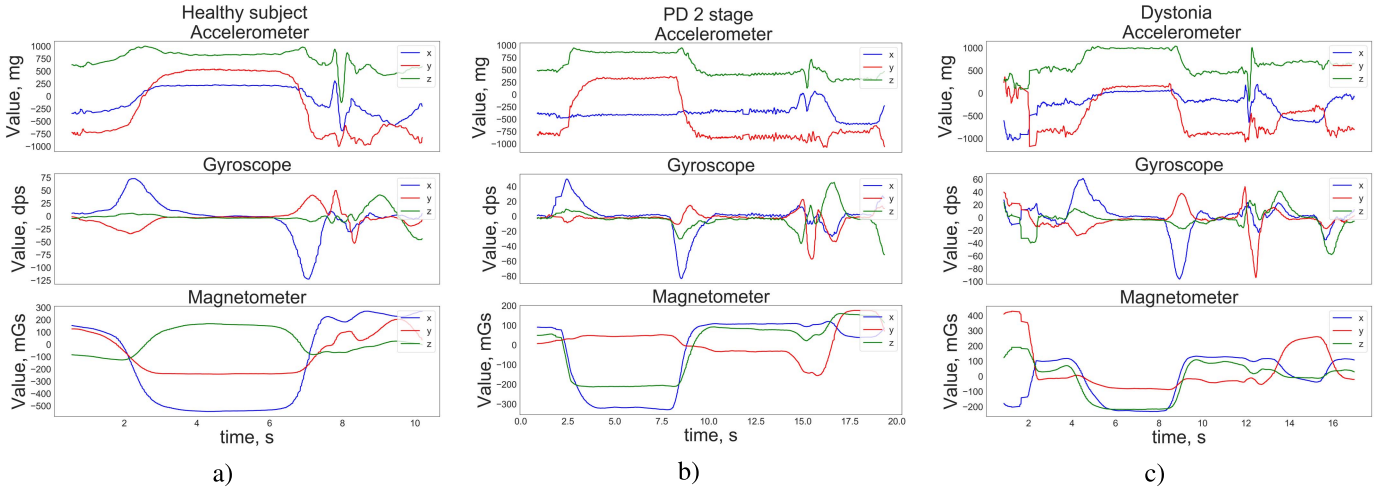

Fig. 2. Sensor data sample for the exercise 7. (a) Healthy subject. (b) Patient with PD (stage 2). (c) Subject with dystonia.

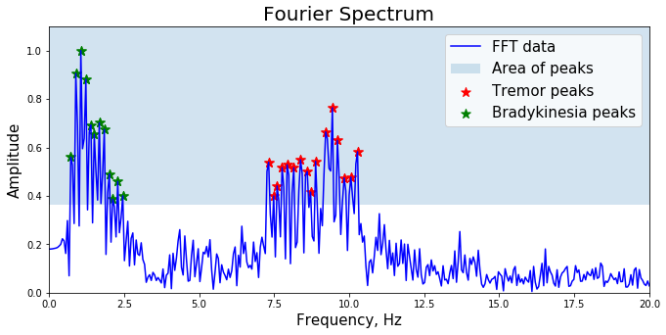

Fig. 3. Fourier spectra of accelerometer signal of a subject with PD doing exercise 7.

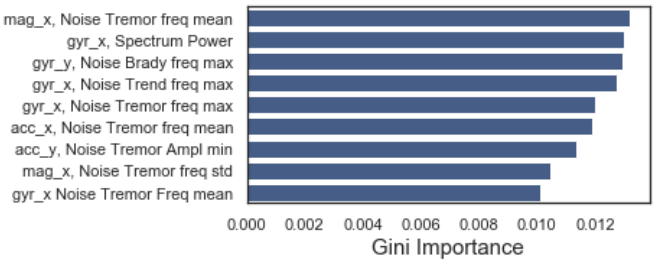

Fig. 4. Features importance.

These components were chosen based on the research reported in [42]. The total ML input size included 42 components: six for each of the PCA, FA, ICA, and MDS. The representations of DR were joined together and moved to the ML model. In summary, when the feature extraction procedure has finished, we applied all the DR methods and formed a new multidimensional (42 new features) feature space by concatenating the output of DR. After that, we applied ML algorithms and selected the best model, which is presented in Section V.

D. ML Algorithms

We used the same ML models for each task as follows.

1) SVM (multi class classification: one versus all approach) [43].

TABLE II
DESCRIPTION OF FEATURES

Name	Description
Standard features*	
mean, std, skew, kurtosis, min, max	Standard statistical features
mean, std	The first derivative of signal
FFT features**	This features were divided into *tremor and bradykinesia* and represent frequency domain.
μ, σ, λ_+ and λ_- of of amplitudes and frequencies of dominant freqs and ampls μ and σ of energy spectrum	Here we understand dominant freqs and ampls as only peak frequencies and amplitudes. Peaks were found as described in The Features Extraction (Section IV-B)

* These features were calculated for each data sample. The derivative were calculated as $Differential[k] = \frac{a_{k+1}-a_k}{t_{k+1}-t_k}$.

** These features were calculated for each segment of FFT. μ, σ λ_-, λ_+ are the mean, std, min and max values

2) Random forest (RF) with different number of estimators and maximal depth [36].
3) Logistic regression (multi class classification: one versus all approach) [44].
4) Naive Bayesian algorithm [45].
5) k-nearest neighbors classifier (with different k) [46].
6) Boosted trees [47].
7) Stacked ensemble model [48].

We did stacking using several base estimators, including SVM (with rbf and linear kernel), RF (with various option of parameters), and logistic regression. Logistic regression proved to be the most efficient in practice and was, therefore, chosen as the final estimator.

Our dataset is an unbalanced one. We used weights to regulate the number of class samples. Since we were analyzing a set of exercises, we combined the features after the DR step. Thus, there were 42 input features for each exercise. For example, if there were three exercises in the set, we performed classification on 126 features. Parameters for the ML algorithms were selected using the grid search approach. All

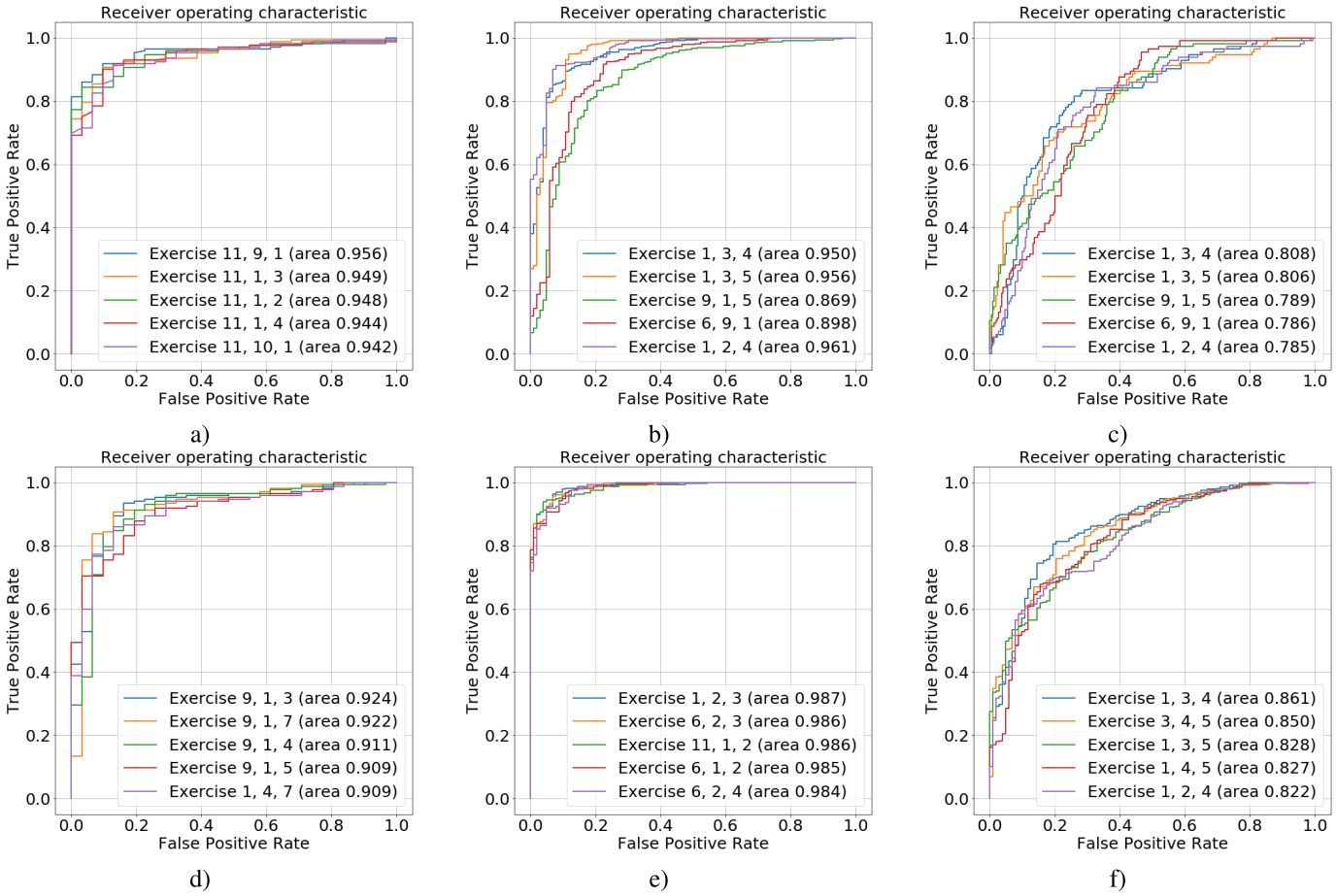

Fig. 5. ROC curves for the best model for cases: (a) diagnosis 1 or 2 stage PD; (b) diagnosis sick or healthy person; and (c) diagnosis PD or other type of neurodegenerative disease. Stacked model for cases: (d) diagnosis 1 or 2 stage PD; (e) diagnosis sick or healthy person; and (f) Diagnosis PD or other type of neurodegenerative.

hyperparameters were selected according to the maximization of the cross validate (CV) metric.

In Section V, we show only the best model for each exercise. In this article, our main goal was not to analyze different models for detecting the PD but to focus on choosing the best set of exercises.

E. Metrics

We applied the standard leave-one-out cross validation (LOO CV) approach for the evaluation of the results. That is, at each step of the CV, patients were divided into train and test parts so that patients from different groups did not overlap and there was only one subject in the test part. We decided to evaluate only the ROC-AUC value as a metric because our dataset is highly imbalanced and the ROC curve is the best option in this case.

V. Results

A. PD Evaluation

The main task was to find the best exercises for performing the analysis of PD. Our idea was that different exercises reflect dissimilar symptoms that are important for various stages and diseases. Therefore, several exercises were used instead of just one. We carried out the experiments on the sets of up to three exercises. Using the brute force approach, we chose the subset of exercises and tried to make the diagnosis of the early stage of PD.

We decided to perform the classification as three different classification problems as follows.

1) Diagnosis 1 or 2 stage PD.
2) Diagnosis sick or healthy person.
3) Diagnosis PD or other types of neurodegenerative disease.

Each task is important for understanding and researching PD and neurodegenerative disease. The earlier PD is detected, the more effective the treatment will be. Therefore, it is necessary to develop ML methods for PD detection. Medications, such as levodopa/carbidopa, help best at stage 1 of PD. There are various therapies, for example, nonpharmacological therapy, that is also most effective at the first stage of PD. Moreover, this can significantly slow down the progress of the disease. The second task gives us the information about general properties of disorders. The second and third ones can be helpful in cases, where the model needs to define a course of treatment.

1) Diagnosis One or Two Stage of PD: The division of the first and second stages of PD is not such a difficult

TABLE III
MODEL BASED ON OPTIMAL EXERCISES

Exercises		Accuracy	Balanced	AUC ROC	f1-score
Important	Additional				
1, 3	2	0.60	0.49	0.82	0.60
1, 3	5	0.62	0.52	0.83	0.62
1, 3	9	0.67	0.56	0.84	0.67
1, 3	11	0.65	0.53	0.86	0.65
1, 3	2, 5	0.59	0.49	0.82	0.59
1, 3	2, 9	0.63	0.51	0.83	0.63
1, 3	2, 11	0.63	0.50	0.85	0.63
1, 3	5, 9	0.65	0.54	0.85	0.65
1, 3	5, 11	0.61	0.50	0.86	0.61
1, 3	9, 11	0.64	0.52	0.86	0.64
1, 3	2, 5, 9	0.65	0.55	0.86	0.63
1, 3	2, 5, 11	0.66	0.54	0.87	0.64
1, 3	2, 9, 11	0.64	0.55	0.87	0.64
1, 3	5, 9, 11	0.65	0.56	0.84	0.63

'Important' means the highly recommended exercises, 'Additional' is Additional exercises. 'Balanced' is a balanced accuracy and AUC ROC is an Area Under the Receiver Operating Characteristic Curve. AUC ROC and f1-score is a micro-averaging of corresponding metrics.

task based on some exercises. Subjects with stage 2 PD have more severe symptoms, and our approach allows us to detect the difference. In this case, exercise 1, exercise 9, and exercise 11 are presented in the best subsets of exercises [see Fig. 5(a) and (d)], so it is highly important for this task.

2) Diagnosis Sick or Healthy Person: General characteristics of the disease are best highlighted in exercise 1, exercise 3, and exercise 2 in Fig. 5(b) and (e). These exercises are more about general movements. Clinicians can use our results for more effective PD detection; moreover, the best set of features is important for researchers for future study.

3) Diagnosis of PD or Other Type of Neurodegenerative Disease: Separation of PD from other types of disorders is a complicated task because different disorders have different symptoms. Moreover, these symptoms have various characteristics depending on the disease. For this task, the exercise 1 and 3 are the best one. The results are presented in Fig. 5(c) and (f).

It is ideal to use all the exercises, but it can be inconvenient and time consuming for the patient and for the doctor. In our case, we highly recommend using exercise 1 and 3, as it has been selected for all problems.

For practical use, it is necessary to build a new model based on a combined group of exercises from each task. We highly recommend using exercise 1 and exercise 3 according to our results. Depending on the formal timing regulations of the diagnosis, we recommend additionally include the following exercises: 2, 5, 9, and 11. A new model based on this group of exercises should be applied for identifying all signs of PD and other neurological disorders. This model will help doctors diagnose the disease.

Finally, we made an RF model for the multinomial classification, i.e., PD (stage 1 and stage 2), other neurological disorders, and healthy subjects. To do this, we have tested all possibilities with highly recommended and additional exercises that are optional features. The results are shown in Table III. It can be observed that the improvement in quality with an increase in the number of exercises is insignificant. Therefore, based on our data, three exercises are sufficient for classification. Moreover, the results of our work state that using ML methods, we can separate any two classes with a probability of about 0.95 (see Fig. 5). However, to build a competitive model for analyzing a multiclass problem, more data are needed and this research is out of the scope of this article.

The proposed solution is a recommendation system (so-called "second opinion" system) helping the doctors make the diagnosis as accurately as possible using the results of this work and avoid potential confusion between the PD stages or between the PD and other neurological disorders during the patient's first visit.

VI. CONCLUSION

This work is one of the first, where the main approach is focused on choosing the best set of exercises for a given diagnosis task. The results showed that exercise 1 does well in all three tasks that were attempted. This is most likely due to the fact that both bradykinesia and rest tremor occur, while the patient is walking and sitting down. Likewise, each task has an exercise that also performs well at differentiating the two classes. This means that there are specific points in differentiating: 1) healthy from early stages of PD; 2) healthy from PD; and 3) PD from other neurodegenerative diseases.

Our results show that based on at least three exercises, we can achieve reasonably good results to detect the disease and the stages of the disease. We have reached 0.9 ROC AUC in each task of the analysis of data from the sensors.

In terms of limitations, the main limitation is connected with wireless communication. Although we have investigated lots of wireless sensing devices and have chosen the most appropriate for our task, sometimes we were losing some data packets. At the same time, the wireless transmitter is the most power-hungry component on board and the battery pack can be depleted in approximately 1 h. Although the battery-based operation is fine in real life it becomes a limiting factor during the experimentation when we need to perform the tasks with a number of subjects in a row and do not know exactly when the battery goes down. In addition, we noticed that wearing even a tiny sensing device might be uncomfortable for the subjects with the fourth stage of the PD.

Our findings reported in this article could be useful for doctors in better understanding what to pay more attention to when evaluating patients. It could be used as a second opinion system to recommend a certain diagnosis.

REFERENCES

[1] R. Cacabelos, "Parkinson's disease: From pathogenesis to pharmacogenomics," *Int. J. Mol. Sci.*, vol. 18, no. 3, p. 551, 2017.
[2] A. Michalski and A. Chwaleba, "Mathematical approach to tremor recognition in extrapyramidal disease," *IEEE Instrum. Meas. Mag.*, vol. 6, no. 1, pp. 57–61, Mar. 2003.
[3] L. Marsili, G. Rizzo, and C. Colosimo, "Diagnostic criteria for Parkinson's disease: From James Parkinson to the concept of prodromal disease," *Frontiers Neurol.*, vol. 9, p. 156, Mar. 2018.
[4] R. Postuma *et al.*, "MDS clinical diagnostic criteria for Parkinson's disease," *Movement Disorders*, vol. 30, no. 12, pp. 1591–1601, 2015.
[5] L. Ferini-Strambi, S. Marelli, A. Galbiati, F. Rinaldi, and E. Giora, "REM sleep behavior disorder (RBD) as a marker of neurodegenerative disorders," *Arch. Italiennes de Biologie-a J. Neurosci.*, vol. 152, nos. 2–3, pp. 129–146, 2014.

[6] R. Agarwal, T. Takeuchi, S. Laroche, and J. Gotman, "Detection of rapid-eye movements in sleep studies," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 8, pp. 1390–1396, Aug. 2005.
[7] B. E. Sakar *et al.*, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 4, pp. 828–834, Jul. 2013.
[8] L. Lonini *et al.*, "Wearable sensors for Parkinson's disease: Which data are worth collecting for training symptom detection models," *npj Digit. Med.*, vol. 1, p. 64, Nov. 2018.
[9] A. H. Butt *et al.*, "Objective and automatic classification of Parkinson disease with leap motion controller," *Biomed. Eng. OnLine*, vol. 17, no. 1, pp. 1–21, Dec. 2018.
[10] E. Kovalenko *et al.*, "Distinguishing between Parkinson's disease and essential tremor through video analytics using machine learning: A pilot study," *IEEE Sensors J.*, vol. 12, no. 10, pp. 11916–11925, May 2020.
[11] A. Bermeo, M. Bravo, M. Huerta, and A. Soto, "A system to monitor tremors in patients with Parkinson's disease," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 5007–5010.
[12] J. Jankovic and E. K. Tan, "Parkinson's disease: Etiopathogenesis and treatment," *J. Neurol., Neurosurg. Psychiatry*, vol. 91, no. 8, pp. 795–808, 2020.
[13] A. Talitckii *et al.*, "Data-driven analysis of Parkinson's disease and its detection at an early stage," in *Proc. 14th EAI Int. Conf. Pervas. Comput. Technol. Healthcare*, New York, NY, USA, May 2020, pp. 419–422.
[14] H. Habibzadeh, K. Dinesh, O. R. Shishvan, A. Boggio-Dandry, G. Sharma, and T. Soyata, "A survey of healthcare Internet of Things (HIoT): A clinical perspective," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 53–71, Jan. 2020.
[15] The National Collaborating Centre for Chronic Conditions, "Diagnosing Parkinson's disease," in *Parkinson's Disease*, B. Higgins, Ed. London, U.K.: Royal College of Physicians, 2006, ch. 5, pp. 29–47.
[16] A. Schrag, L. Horsfall, K. Walters, A. Noyce, and I. Petersen, "Prediagnostic presentations of Parkinson's disease in primary care: A case-control study," *Lancet Neurol.*, vol. 14, no. 1, pp. 57–64, Jan. 2015.
[17] J. Levin, A. Kurz, T. Arzberger, A. Giese, and G. U. Höglinger, "The differential diagnosis and treatment of atypical parkinsonism," *Deutsches Aerzteblatt Online*, vol. 113, pp. 61–69, Feb. 2016.
[18] E. D. Louis, "Essential tremor," *Lancet Neurol.*, vol. 4, no. 2, pp. 100–110, 2005.
[19] G. Gugliandolo *et al.*, "A movement monitoring system for patients of neurodegenerative diseases," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2018, pp. 1–6.
[20] H. Ali Hashim, S. L. Mohammed, and S. K. Gharghan, "Accurate fall detection for patients with Parkinson's disease based on a data event algorithm and wireless sensor nodes," *Measurement*, vol. 156, May 2020, Art. no. 107573.
[21] B. Andò *et al.*, "A wearable device to support the pull test for postural instability assessment in Parkinson's disease," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 1, pp. 218–228, Jan. 2018.
[22] G. Gugliandolo *et al.*, "A movement-tremors recorder for patients of neurodegenerative diseases," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 5, pp. 1451–1457, May 2019.
[23] N. Shawen *et al.*, "Role of data measurement characteristics in the accurate detection of Parkinson's disease symptoms using wearable sensors," *J. Neuroeng. Rehabil.*, vol. 17, p. 52, Apr. 2020.
[24] M. Heijmans, J. Habets, M. Kuijf, P. Kubben, and C. Herff, "Evaluation of Parkinson's disease at home: Predicting tremor from wearable sensors," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 584–587.
[25] P. Pierleoni, A. Belli, O. Bazgir, L. Maurizi, M. Paniccia, and L. Palma, "A smart inertial system for 24h monitoring and classification of tremor and freezing of gait in Parkinson's disease," *IEEE Sensors J.*, vol. 19, no. 23, pp. 11612–11623, Dec. 2019.
[26] B. Andò *et al.*, "A measurement system to monitor postural behavior: Strategy assessment and classification rating," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 8020–8031, Oct. 2020.
[27] B. Mariani, M. C. Jiménez, F. J. G. Vingerhoets, and K. Aminian, "On-shoe wearable sensors for gait and turning assessment of patients with Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 155–158, Jan. 2013.
[28] M. Ferrarin, I. Carpinella, M. Rabuffetti, M. Rizzone, L. Lopiano, and P. Crenna, "Unilateral and bilateral subthalamic nucleus stimulation in Parkinson's disease: Effects on EMG signals of lower limb muscles during walking," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 2, pp. 182–189, Jun. 2007.
[29] M. Nilashi, O. Ibrahim, S. Samad, H. Ahmadi, L. Shahmoradi, and E. Akbari, "An analytical method for measuring the Parkinson's disease progression: A case on a Parkinson's telemonitoring dataset," *Measurement*, vol. 136, pp. 545–557, Mar. 2019.
[30] S. Lahmiri, "Gait nonlinear patterns related to Parkinson's disease and age," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 7, pp. 2545–2551, Jul. 2019.
[31] S. Barrantes *et al.*, "Differential diagnosis between Parkinson's disease and essential tremor using the smartphone's accelerometer," *PLoS ONE*, vol. 12, no. 8, Aug. 2017, Art. no. e0183843.
[32] P. Tahafchi and J. W. Judy, "Freezing-of-Gait detection using wearable-sensor technology and neural-network classifier," in *Proc. IEEE SENSORS*, Oct. 2019, pp. 1–4.
[33] P. Locatelli and D. Alimonti, "Differentiating essential tremor and Parkinson's disease using a wearable sensor—A pilot study," in *Proc. 7th IEEE Int. Workshop Adv. Sensors Interfaces (IWASI)*, Jun. 2017, pp. 213–218.
[34] M. A. Lones, J. E. Alty, P. Duggan-Carter, A. J. Turner, D. R. S. Jamieson, and S. L. Smith, "Classification and characterisation of movement patterns during levodopa therapy for parkinson's disease," in *Proc. Companion Publication Annu. Conf. Genetic Evol. Comput.*, Jul. 2014, pp. 1321–1328.
[35] S. Patel *et al.*, "Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 6, pp. 864–873, Nov. 2009.
[36] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[37] B. Schölkopf, A. J. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Proc. ICANN*, 1997, pp. 583–588.
[38] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.
[39] R. B. Darlington, *Factor Analysis*. 2008.
[40] D. B. Skillicorn, *Independent Component Analysis (ICA)*. New York, NY, USA: Chapman & Hall, 2007.
[41] H. T. Shen, "Multidimensional scaling," in *Encyclopedia of Database Systems*. Boston, MA, USA: Springer, 2009.
[42] S. Patel *et al.*, "Analysis of the severity of dyskinesia in patients with Parkinson's disease via wearable sensors," in *Proc. Int. Workshop Wearable Implantable Body Sensor Netw. (BSN)*, Apr. 2006, p. 126.
[43] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory (COLT)*, 1992, pp. 144–152.
[44] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*. New York, NY, USA: Wiley, 1989.
[45] I. Rish, "An empirical study of the naive bayes classifier," IBM Res. Division, Thomas J. Watson Res. Center, Yorktown Heights, NY, USA, Tech. Rep. RC 22230 (W0111-014), 2001.
[46] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy K-nearest neighbor algorithm," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 4, pp. 580–585, Aug. 1985.
[47] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.
[48] F. Gunes, R. Wolfinger, and P. Y. Tan, "Stacked ensemble models for improved prediction accuracy," in *Proc. SAS Global Forum*, Orlando, FL, USA, 2017, pp. 1–19.

Aleksandr Talitckii received the bachelor's degree in high energy physics from the Moscow Institute of Physics and Technology, Dolgoprudny, Russia, in 2018, and the M.Sc. degree in information science and technology from the Skolkovo Institute of Science and Technology, Moscow, Russia, in 2020. He is currently pursuing the Ph.D. degree with the Arizona State University, Tempe, AZ, USA.

In 2019, he was an Intern with the Fondazione Bruno Kessler Research Center, Italy. His research interest includes data analysis in medical-related areas.

Anna Anikina received the bachelor's degree in applied physics and mathematics from the Moscow Institute of Physics and Technology, Dolgoprudny, Russia, in 2018, and the M.Sc. degree in information science and technology from the Skolkovo Institute of Science and Technology, Moscow, Russia, in 2020.

She is currently a Research Engineer with Skoltech. Her research interests cover machine learning and medical applications.

Ekaterina Kovalenko received the B.A. degree in mathematics from the University of California at Berkeley, Berkeley, CA, USA, and the M.S. degree in data science from the Skolkovo Institute of Science and Technology, Moscow, Russia, in 2020, where she is currently pursuing the Ph.D. degree in data science.

Her research areas include machine vision, signal processing, and machine learning in medical-related fields.

Aleksei Shcherbak received the M.Sc. degree in physics from Moscow State University, Moscow, Russia, in 2019. He is currently pursuing the Ph.D. degree with the Skolkovo Institute of Science and Technology, Russia.

His research interests include wearable sensing, machine learning, and medical-related applications.

Oscar Mayora received the Ph.D. degree from the University of Genoa, Genoa, Italy.

He joined the Advance Interactive Systems Laboratory, VTT Electronics, Oulu, Finland, as an European Research Consortium for Informatics and Mathematics (ERCIM) Visiting Research Fellow. In 2001, he was appointed as an Associate Professor and the Head of the Graduate Program in computer science at Tecnológico de Monterrey, Monterrey, Mexico. From 2004 to 2016, he was the Head of the Ubihealth area at Create-Net, Trentino, Italy. He is currently the Head of the Digital Health Unit, FBK Health and Wellbeing Center, Trentino, a Scientific Coordinator of the EIP-AHA Trentino Reference Site, Trentino, and responsible for the coordination of the project of TrentinoSalute4.0 initiative. He has authored or coauthored over 150 scientific papers and has extensively participated and coordinated research and innovation projects in the field of pervasive health technologies.

Dr. Mayora was the former President of ACM Special Interest Group on Computer–Human Interaction (SIG-CHI) for Mexico. He is the Founder of the International Pervasive Health Conference.

Olga Zimniakova received the M.D. and Ph.D. degrees.

She is currently a Docent from the Biomedical University of Innovation and Continuing Education, Moscow, Russia, and a Neurologist from the Federal Neurological Center, Extrapyramidal Disorders and Mental Health Federal Medical and Biological Agency, A. I. Burnazyan Federal Medical and Biophysical Center, Moscow.

Ekaterina Bril received the M.D. and Ph.D. degrees.

She is currently the Head of Department with the Federal Neurological Center of Extrapyramidal Disorders and Mental Health Federal Medical and Biological Agency, A. I. Burnazyan Federal Medical and Biophysical Center, Moscow, Russia.

Maxim Semenov received the M.D. and Ph.D. degrees.

He is currently the Head of the Neurosurgical Department, the Head of the Center for Neurosurgery and Vascular Neurology, and a Neurologist with the A. I. Burnazyan Federal Medical and Biophysical Center, Moscow, Russia.

Dmitry V. Dylov (Member, IEEE) received the M.Sc. degree in applied physics and mathematics from the Moscow Institute of Physics and Technology, Moscow, Russia, in 2006, and the Ph.D. degree in electrical engineering with Princeton University, Princeton, NJ, USA, in 2010.

He is currently an Associate Professor with the Center for Computational and Data-Intensive Science and Engineering, Skolkovo Institute of Science and Technology, Moscow, where he is also the Head of the Computational Imaging Group.

Andrey Somov received the Ph.D. degree in information and communication technology and the Diploma degree in electronics engineering from the MATI-Russian State Technological University, Moscow, Russia, in 2004 and 2006, respectively, and the Ph.D. degree in the field of power management in wireless sensor networks (WSNs) from the University of Trento, Trento, Italy, in 2009.

He was a Senior Researcher with the Center of REsearch And Telecommunication Experimentations for NETworked communities, Trento, Italy, from 2010 to 2015, and a Research Fellow with the University of Exeter, Exeter, U.K., from 2016 to 2017. He is currently an Assistant Professor with the Skolkovo Institute of Science and Technology, Moscow. He has authored or coauthored more than 80 papers in peer-reviewed international journals and conference proceedings. His current research interests include wearable sensing, machine learning, cognitive IoT, and associated proof-of-concept implementation.

Dr. Somov holds some awards in the fields of WSN and IoT, including the Google IoT Technology Research Award in 2016 and the Best Paper Award at the IEEE IoP conference in 2019.