Dominant Hand Invariant Parkinson's Disease Detection Using 1-D CNN Model and STFT-based IMU Data Fusion

Aleksei Shcherbak[1], Ekaterina Kovalenko[2], Ekaterina Bril[3,1], Anna Baldycheva[4], Andrey Somov[1]
[1]*Skolkovo Institute of Science and Technology, Moscow, Russia*
Aleksei.Shcherbak@skoltech.ru
[2]*No affiliation*
[3]*Burnazyan Federal Medical and Biophysical Center, Russia*
[4]*University of Exeter, UK*

2023 IEEE 32nd International Symposium on Industrial Electronics (ISIE) | 979-8-3503-9971-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISIE51358.2023.10228119

***Abstract*—In this paper, we report on a dominant hand invariant 1-D convolution model for distinguishing between Parkinson's disease (PD) and healthy subjects. For realizing this approach, we propose an STFT-based method for IMU data axes fusion. We learn how both the different frequency ranges and the period of the STFT window affect the efficiency of Parkinson's disease detection. We test this solution on the dataset collected from 58 subjects. Our results show superiority of the proposed axes fusion method in 70% of cases in comparison with the state-of-the-art. Results also prove the efficiency of the proposed 1-D convolution hand invariant model with the best scores 98% of AUC and 92% of F1 and accuracy metrics. In addition, we show the STFT window must be at least 2 seconds of length, while the frequency range must include the frequencies below 3 Hz.**

***Index Terms*—Parkinson's disease, STFT, CNN, IMU**

I. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder becoming one of the widespread neurological diseases [1]. Its world-wide burden increased from 2.5 million patients in 1990 to 6.1 millions in 2016 [2] [3]. The increasing number of PD patients expected to double by 2050 [3] [4].

The PD is typically characterised and can be detected through the visual signs of specific movement disorders, e.g. tremor, which can be characterized as involuntary oscillations of human extremities and is the most prominent [5]. From the clinical point of view, there are no unambiguous biomedical or radiological biomarkers reliably identifying the PD [6]. PD diagnostics is, therefore, relied on the clinical observations. What is worse, it is impossible to identify the PD during the lifetime, and only 75-95% of PD diagnosis are confirmed after the autopsy [7]. The PD diagnostics has a probabilistic character with up to 25% of misdiagnosis [7]. To recognise the PD, clinical observations may be conducted in the compliance with Movement Disorder Society (MDS) Unified Parkinson's Disease Rating Scale (MDS-UPDRS), which is a commonly used tool to assess the PD stage. However, MDS-UPDRS does not define the PD unlike the MDS, where PD is defined as a kind of parkinsonism fulfilled with the special MDS-PD criteria [7]. According to MDS, the patients diagnosed with the PD have to satisfy the parkinsonism criteria. It requires bradykinesia with a combination of the rigidity or/and rest tremor during clinical observations, and meet the MDS-PD diagnostic criteria including two distinct levels: (i) clinically established PD maximizing the specificity with a value of at least 90%, and (ii) clinically probable PD, balancing sensitivity and specificity with a goal to make both values of at least 80%. Although the rest tremor is considered as a cardinal symptom of PD, the patients may suffer from the action tremor as well, which comprises postural, kinetic and intention tremor [8]. Action tremor in pure or mixed form (with rest tremor) has up to 92% of prevalence [9]. The special case of postural tremor is a tremor in outstretched arms position, so-called re-emerged tremor, which may be considered as a type of the rest tremor, because it shows the same frequencies and responsive to dopaminergic therapy [10].

Apart from the tremor diversity and probability character of the PD diagnosis, there are diseases with similar or overlapping symptoms, e.g. essential tremor, which may complicate the PD diagnosis even more [11], especially when the signs of different diseases are present simultaneously [8].

The problem is that the clinical PD diagnostics has probabilistic and often controversial symptoms and still does not have a clearly defined diagnosis criteria even for distinguishing healthy subject from the PD ones. To sum up, the diagnosis usually relies on the MDS formally defining the PD, MDS-UPDRS test and observation of bradykinesia, rigidity and various kinds of tremor, and is characterized by up to 25% of misdiagnosis.

To address this problem, the recent trend is the application of Machine Learning (ML) for the approximation of complicated nonlinear dependencies to overcome the PD misdiagnosis problem. In this context, research works report on the investigation of the spiral-based methods [12], video processing [13], the keyboard pressings analysis [14], voice analysis methods [15] - all for detecting the PD. However,

This research was funded through the RSCF grant no. 23-21-00473.

the popular approach for the PD detection includes Inertial Measurement Unit (IMU) based ML data analysis where IMU is used in a shape of wearable sensors [16] [17] as well as smartphones/smartwatches [18]. The lion share of IMU-based methods is applied to the gait analysis, while the second place is kept by the tremor analysis [19]. To both of them, classical ML models and neural networks are applied. However, to the best of our knowledge, the most of tremor related research works are relied on classical ML models [16] while deep-learning models are common for the gait-analysis [20] [21] which is additionally confirmed by [22].

One of recent works relies on Short-Time Fourier Transform (STFT) for the data analysis [18], where the authors use standard statistical properties of the tremor spectrum from each of three IMU axes as features. Authors use Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive classifiers and Artificial Neural Network (ANN) consisting of 2 linear layers with 5 neurons each. ANN results are the worst among the tested classifiers. The next example uses STFT and Convolutional Neural Network (CNN) [23], where the authors applied 1-D convolution to the IMU data. The authors reported on successful results with an average accuracy of around 95%. However, the kernel size was 3, and the goal was to recognize a human activity for clinical balance assessment, which is not directly related to the PD detection. In [17] LeNet-based CNN was proposed for hand tremor detection on the 10 subjects dataset featured by 97% of accuracy, outperforming SVM and Back Propagation (BP) models.

All these works have disadvantages as they are dominant hand depended and some of them are not invariant with respect to sensors orientation. In addition, STFT window length choice, as a rule, is not explained enough, while IMU data fusion way produces parasitic harmonics.

In this work, we address these disadvantages. In particular, we apply the 1-D CNN model to the IMU-based PD tremor analysis in the frequency domain using the proposed STFT-based IMU data fusion method. Achieved results show efficiency of our model with up to 98% of AUC and 92% of F1 and accuracy metrics, and superiority of proposed IMU data fusion method in 70% of cases. Using the proposed model as an assessment tool, we learnt the STFT window must be at least 2 seconds of length, while the frequency range must go below 3 Hz.

II. Methods

A. Methodology

This research was conducted in collaboration with Burnazyan Federal Medical and Biophysical Center (Russia) where data from the patients were collected. Information about the patients includes two parts: the experiment data and medical conclusion from neurologists, who examined the patients before experiments, including diagnostics date, diagnosis, Hoehn-Yahr stage of the disease, age, etc. All patients were assigned with their own ID for distinction purposes, and data was saved anonymously. The most of diagnosis and stages were clear defined, however, there were cases of uncertainty, which led to sending patients for additional examinations. In such situations, clearly defined clinical diagnoses came later and expressed a group opinion of several neurologists (3-5 persons), relying on additional comprehensive examinations. Before the experiments, patients were given a description, what is going on, what time it takes and how the collected data is used. Then they signed the agreement to participate in this study. They were also notified about the video record of the experiments. The last stage of the data collection was the experiments consisting of 11 exercises enumerated in Table I.

TABLE I
Exercises Summary

Exercise Number	Exercise Description
1	Stand up from the chair, walk there and back, and sit down again
2	Stand, straighten the arms and do pronation/supination of the forearms
3	Sit down, drop the arms and rest
4	Bend the arms and open/close thumb and index fingers, while the rest fingers are held closed to the palm
5	Straight the arms and then alternately touch the nose by the index fingers of them
6	Outstretched arms position
7	Tight the nut on the anchor
8	Fill the glass of water form the bottle, bring the glass to the mouth, keep this position 3 seconds and put the glass down on the table again
9	Hands are at the table, clench and unclench the fist
10	Stand, bend the arms and align them within the line at the height of the breast, keeping the palms facing the breast
11	Stand, bend the arms and align them within the line at the height of the breast, keeping the palms facing forward

TABLE II
Patients Demographic Distribution

Patient Type	Male	Female	Total
PD Stage 1-1.5	3	5	8
PD Stage 2-2.5	11	9	20
PD Stage 3	2	2	4
PD in Total	16	16	32
H in Total	20	6	26

Some of these exercises were adopted from MDS-UPDRS assessment scale, while others are used by neurologists at the medical center during clinical observations. Before each exercise, it was explained and visually demonstrated how to perform the upcoming task. In total, the dataset was collected from 32 PD patients within the ages from 55 to 84 years. Among them, there are 8 people with 1-1.5 PD stage (3 patients with 1.5 stage), 20 peoples with the 2nd PD stage (3 of them with 2.5 stage) and 4 people with the 3rd stage of PD. The gender distribution is equal. Additionally, data was collected from 26 healthy subjects (H) to provide the dataset with a control group. Part of them are within the same age range, while the others are younger, from 22 to 45 years.

B. Experimental Testbed

One of the key features of this work is the simultaneous application of two sensors placed on both hands and synchronized with 1 millisecond accuracy. The SensorTile sensor node [24] was used as the hardware platform for such sensors, because it has Bluetooth LE communication abilities, microSD slot, several IMU chips, it is based on STM32L476JG with reprogramming opportunity, and lightweight, compact and easy to use with a bracelet. Regarding Bluetooth and sensor data collection/transmission, there is a well-known problem of packet loss. It appears because the human body absorbs 2.4 GHz Bluetooth LE signal from closely attached sensors very well. It leads to high signal attenuation. To overcome this problem and provide data synchronization in time, we reprogrammed SensorTile, developed a separate synchronization module and Android application to manage the data collection. The half-period of a minimum sampling frequency for tremor investigation (40 Hz frequency and 12.5 milliseconds, respectively) was taken as the acceptable synchronization error since the PD tremor frequency above 20 Hz was not seen neither in relevant literature nor in our experiments. As a result, each sensor collects the accelerometer and gyroscope data from the on-board LSM6DSM IMU chip with a frequency of 416 Hz and resolution of ± 2 g and ± 250 dps, respectively. Data is saved to the on-board microSD card into a txt file formatting as a csv file to easy visual error/data consistency inspection after experiments. Each data row in the file contains the IMU data and its time-stamp with a 30.6 microseconds resolution, since time-stamps were derived from the STM32L476JG RTC block clocked by external 32768 Hz oscillator. During experiments, the bracelets with sensors were put on the wrist of patient's hands tightly but comfortably to prevent the unwanted parasitic mechanical oscillations.

C. Data Preprocessing and Feature Engineering

Experimental setup produced synchronized IMU data from two sensors with a frequency of 416 Hz and no missing values, thus no data restoration or interpolation was applied except for the time-stamp based data aligning. STFT with a Blackman-Harris window was used to convert the IMU signal from the time domain to the frequency domain and represent spectrum change over time for each exercise. The low-pass filter to remove gravitational component from the raw time-series data was not used, because in the frequency domain, after STFT transformation, near-zero components can be removed just by not including them into the input data for the machine learning model, what was done.

Wearable sensors placement, its orientation and way of doing exercises inevitably vary from patient to patient. It significantly reduces data uniformity and can negatively affects efficiency of the machine learning model. It is reasonable to make input data invariant with respect to sensors orientation to mitigate this problem. For this purpose, we combined data from 3 separate axes of IMU into the single one using equation (2), where STFT denotes the absolute value of STFT, S_j is the resulting spectrum, $a_{x/y/z}$ are corresponding acceleration components, j is a frequency index, i is a sample index and N is the STFT window length in samples number. To make the same thing, in [17] authors used equation (1), however this equation gives harmonics unlike the proposed equation (2). It can be easily observed via calculations replacing a_i with $sin(x)$ or $cos(x)$ in (1). These parasitic harmonics may reduce the efficiency of machine learning in most cases, and it is shown in Section III. Furthermore, unlike (1), proposed equation considers not only amplitude oscillations of the resulting vector, but its angle oscillations as well. It allows us to consider rotational components of the resulting vector in xyz coordinates, which is important especially when amplitude is constant or oscillations are primarily rotational.

$$S_j = \mathrm{STFT}_j \left(\left\{ \sqrt{a_{x,i}^2 + a_{y,i}^2 + a_{z,i}^2} \right\}_{i=0}^{N-1} \right) \tag{1}$$

$$S_j = \Big[\mathrm{STFT}_j^2(\{a_{x,i}\}_{i=0}^{N-1}) + \\ + \mathrm{STFT}_j^2(\{a_{y,i}\}_{i=0}^{N-1}) + \mathrm{STFT}_j^2(\{a_{z,i}\}_{i=0}^{N-1}) \Big]^{1/2} \tag{2}$$

It is known [18] and was observed during experiments that tremor by its nature may not be present during all time of an exercise and its amplitude may vary [25]. Moreover, in some exercises, e.g. 5,6,8 (see Table I) tremor is present in some specific moments only, when it was provoked. It was noticed for re-emerged tremor of outstretched arms position that it may commence after a variable delay up to several seconds [8]. Additionally, in research [26] author noticed as well, that tremor specific harmonicas and tremor amplitude variations over time cannot be neglected. In this way, it looks reasonable to pass signal to the machine learning model from the entire exercise within all predefined frequency range to analyze harmonics and cumulatively absorb frequency features rather than slice exercise into short windows, extract only statistical features of dominant frequency and consider them as a one unit of a train/test sample as it is doing in various papers [18], because in this case PD-marked slices may not contain PD-related frequency features and do not consider harmonics, which can reduce feature comprehensiveness of the model. Following this idea, in this study, in each exercise STFT was applied to the whole time-series signal with a duration of the assumed duration of the active phase of an exercise. The length of STFT/FFT windows defines the frequency resolution of STFT/FFT transformations, and this value varies in relevant publications. More duration - more frequency resolution, but less time resolution, and it is not clear what is the optimal choice for PD detection. Several lengths of STFT windows were tested in this research.

The final step of the data preprocessing was the spectrum transformation. Three options were tested: no spectrum transformation, square of the spectrum and logarithm from the spectrum. The last one demonstrated the best results, so it was used in pipelines for the final results in Section III. Remarkably, this result is in accordance with the Weber-Fechner law of psychophysics [25] [27], which predicted the logarithmic relationship between tremor amplitude and tremor severity.

The next challenge of the PD-related machine learning research is the different manifestation of the tremor in both hands. Bilateral symmetric parkinsonism considered as an excluding criteria for PD diagnosis [7], and some papers claim to use IMU data from the dominant hand only [17]. It requires not just conduct standardized experiments, but also do it in different ways depending on the dominant hand, and twice, if both hands have tremor (later PD stages or ET) and only one sensor is used. In this research, sensors from both hands are used and synchronized, which opens the way to make the machine learning model invariant with respect to the dominant hand and simplifies experimental part or diagnosis part for neurologists in practical use.

D. Models

During preprocessing, IMU data for each exercise were divided into *N* number of time-slots to which the STFT with a fixed window size was applied. Preprocessing for each patient and each exercise produces the output data with a size of $N \times K$, where *K* is the number of STFT frequencies used in the final preprocessing result. *K* also depends on the STFT window size, so $K = (F_{top} - F_{bot}) * W_{size} + 1$, where F_{bot}[Hz] is the bottom frequency, F_{top}[Hz] is the top frequency, W_{size}[Seconds] is the STFT window size. In relevant papers, STFT windows overlapping is a fixed value and commonly mentioned as the STFT characteristic. In this paper, the role of fixed values was given to the STFT window size and the N number of its output points, while overlapping is a various value and depends on both of them and duration of an exercise (its active phase to be processed). As a result, after the data prepossessing, there are N the STFT output points, representing the spectrum for different time moments during an exercise from the beginning to the end. The idea of proposed neural network is to apply N 1-D convolutions to the IMU signal from the entire exercise (its active phase) in frequency domain and get the results, proportional to the degree of similarity between IMU spectrum and the corresponding convolutions coefficients, representing the spectral features, automatically found during the model training process. Spectrum features expressed in convolutions weights are automatically found during the training process, therefore no manual splitting of IMU spectrum into several specific regions, typical for different kind of tremor, and their analysis are required. Model contains two identical sets of N 1-D convolutions (kernal size = K), the first one is for the left hand, and the second one is for the right hand, followed by the 1-D maxpool layer, playing the role of automatic selection of the dominant hand. Output of the model is finalised by the sigmoid layer for the binary classification.

III. Results

Proposed above neural network model was tested with different STFT window sizes, variable number of output points, several frequency ranges, and various spectrum transformations. Tested STFT windows were with a length of 1, 2 and 5 seconds, tested numbers of output points is 20, 50 and 100 points, tested frequency ranges were between 1 Hz and 20 Hz. As it was explained in Section II-C, only logarithm spectrum transformation was used for the final results.

During experiments, not all patients were able to perform all exercises, therefore the number of patients is different in each exercise. To discard the problem of class imbalance and improve the efficiency of model training, the numbers of PD and *H* subjects in training datasets were the same and equaled to the minimum number of patients among PD and *H* group of the exercise to be used for the training. Distribution of PD and *H* subjects for test datasets was equal as well to simplify metrics calculations and ensure applicability of some of them. All final results was derived from cross-validation conducted 5 times (epochs) representing the mixture of Leave-one-subject-out (LOSO) and k-fold: 2 patients from each of PD an *H* groups were left for the test, while the rest patients were used for the training process. 4 patients in total were used for tests while the number of training patients was variable depending on an exercise. Final scores represented as a mean value taken from the 5 cross-validation epochs $\pm 95\%$ confidence interval. As a rule, number of PD patients was higher than *H*, therefore the overall size of train+test datasets was limited by the double number of *H* patients, and not all PD subjects were included into datasets in each cross-validation epoch. To solve this problem, before starting of each cross-validation epoch, patients from PD and *H* lists were randomly chosen to form train/test datasets. Thus, the final scores cover almost all patients from PD group in each exercise, not only those, who were included in the training/test datasets in some certain cross-validation epoch.

Prior to starting the training, data normalization was applied. The simplified shape of the dataset at the input of the model is $(P \times N \times K)$, where P is the number of patients. The standard scaler was applied to the training dataset for each N and K within all P, so each point of $N \times K$ array of each patient where scaled in a way to make the zero mean and unit standard deviation (std) of them among all patients in the training dataset. Normalization in this exact way gives the best result for our model. This procedure was conducted each time before each iteration of cross-validation inside the epoch. Mean and std coefficients were saved and then applied to the test data, thus patients from the test dataset did not affect the training dataset normalization.

In all cases, training process consisted in 90 epochs, Adam optimizer was applied with a learning rate of $1e^{-3}$, Binary Cross Entropy was used as a loss function, no validation was used to maximize the useful dataset size and batch size was equal to all train dataset. Several metrics, such as Area Under the Curve (AUC), F1 score, Precision, Recall, Specificity and Accuracy, were used to rate the ML model performance to make our results easy to compare, they were calculated as follows:

TABLE III
COMPARISON STUDY FOR TYPICAL EQUATION (1) AND PROPOSED EQUATION (2) FOR IMU DATA AXES FUSION

Exercise	Equation	AUC	F1 binary	Precision	Recall	Specificity	Accuracy
1	Proposed method	0.84 ±0.02	0.78 ±0.04	0.74 ±0.01	**0.83 ±0.08**	0.70 ±0.04	**0.77 ±0.02**
	Typical way	0.86 ±0.03	0.80 ±0.0	0.79 ±0.04	0.77 ±0.03	0.83 ±0.04	0.76 ±0.04
2	Proposed method	**0.91 ±0.02**	**0.82 ±0.01**	**0.77 ±0.02**	**0.89 ±0.01**	**0.73 ±0.03**	**0.81 ±0.02**
	Typical way	0.86 ±0.03	0.79 ±0.02	0.73 ±0.03	0.87 ±0.05	0.68 ±0.06	0.78 ±0.03
3	Proposed method	**0.84 ±0.02**	**0.77 ±0.02**	0.72 ±0.02	**0.83 ±0.04**	0.67 ±0.02	**0.75 ±0.02**
	Typical way	0.81 ±0.01	0.76 ±0.04	0.73 ±0.03	0.80 ±0.05	0.70 ±0.03	**0.75 ±0.04**
4	Proposed method	**0.90 ±0.01**	**0.81 ±0.01**	**0.78 ±0.02**	0.84 ±0.02	**0.76 ±0.03**	**0.80 ±0.01**
	Typical way	0.89 ±0.01	0.78 ±0.02	0.72 ±0.02	0.86 ±0.02	0.66 ±0.03	0.76 ±0.02
5	Proposed method	**0.89 ±0.02**	**0.85 ±0.02**	0.81 ±0.01	**0.88 ±0.03**	0.80 ±0.02	**0.84 ±0.02**
	Typical way	0.87 ±0.01	**0.85 ±0.02**	0.83 ±0.04	0.86 ±0.02	0.82 ±0.04	**0.84 ±0.03**
6	Proposed method	**0.98 ±0.00**	**0.92 ±0.03**	**0.94 ±0.02**	**0.90 ±0.04**	**0.95 ±0.02**	**0.92 ±0.03**
	Typical way	0.94 ±0.01	0.86 ±0.01	0.84 ±0.03	0.88 ±0.01	0.83 ±0.04	0.85 ±0.02
7	Proposed method	**0.63 ±0.04**	**0.63 ±0.04**	**0.62 ±0.04**	**0.64 ±0.06**	**0.60 ±0.05**	**0.62 ±0.04**
	Typical way	0.55 ±0.03	0.56 ±0.03	0.55 ±0.03	0.57 ±0.04	0.53 ±0.07	0.55 ±0.03
8	Proposed method	**0.87 ±0.01**	**0.84 ±0.00**	**0.83 ±0.01**	**0.86 ±0.02**	**0.82 ±0.02**	**0.84 ±0.00**
	Typical way	0.82 ±0.02	0.78 ±0.02	0.73 ±0.02	0.84 ±0.03	0.69 ±0.03	0.76 ±0.02
9	Proposed method	0.82 ±0.02	**0.77 ±0.03**	0.69 ±0.03	**0.87 ±0.03**	0.60 ±0.04	**0.74 ±0.03**
	Typical way	0.83 ±0.04	0.74 ±0.05	0.71 ±0.05	0.77 ±0.06	0.68 ±0.06	0.72 ±0.06
10	Proposed method	0.82 ±0.03	0.73 ±0.05	0.74 ±0.06	0.72 ±0.04	**0.74 ±0.07**	0.73 ±0.05
	Typical way	0.85 ±0.03	0.76 ±0.04	0.75 ±0.05	0.78 ±0.05	0.73 ±0.05	0.76 ±0.05
11	Proposed method	**0.82 ±0.02**	**0.72 ±0.04**	**0.72 ±0.05**	0.73 ±0.03	**0.71 ±0.06**	**0.72 ±0.04**
	Typical way	0.76 ±0.04	0.69 ±0.04	0.61 ±0.04	0.78 ±0.05	0.50 ±0.05	0.64 ±0.05
AVG	Proposed method	**0.85**	**0.78**	**0.76**	**0.82**	**0.74**	**0.77**
	Typical way	0.82	0.76	0.72	0.80	0.69	0.75

$$Precision = \frac{TP}{TP+FP}; \quad Specificity = \frac{TN}{TN+FP}; \quad (3)$$

$$Recall = \frac{TP}{TP+FN}; \quad Accuracy = \frac{TP+TN}{TP+TN+FP+FN}; \quad (4)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

where TN/TP is the true negative/positive rate, and FN/FP is the false negative/positive rate. Three series of experiments were conducted. During all of them, the duration of the active phase of each exercise was equal to 30 seconds, except for 1, 5, 7 and 8 exercises, where the duration was 15, 15, 10 and 20 seconds, respectively. In the first series of experiments, different lengths of STFT window were tested: 1, 2 and 5 seconds with 50 output points each and 1-20 Hz frequency range. In average, clear, but insignificant difference was between 1 second and 2 seconds of STFT window length in favour of the last one, while the difference between 2 seconds and 5 seconds was tiny and not in favor of 5 seconds, so the best result was achieved with 2 seconds of STFT window length for 50 output points, and this length was used for next experiments. One of them was devoted to identify, what is the optimal number of STFT output points for preprocessing. 20, 50 and 100 numbers were tested with 2 seconds STFT window length and 1-20 Hz frequency range. Almost no any difference was noticed between them, just average value, excluding 7 exercise (this exercise does not work), was growing a little with increasing the number of output points, but the computational cost was much higher. Thus, the increasing the number of output points is a pointless thing. The next step could be search the minimum number of output points, however, 20 points is already closed to the minimum possible value for 30 seconds of active phase interval of an exercise (16 points at 2 seconds STFT window length), therefore 20 points is considered as the optimal value in our constraints. The third series of experiments was about testing different frequency ranges within 1-20 Hz interval. 1-20 Hz, 1-12 Hz, 1-8 Hz and 3-8 Hz ranges were tested to figure out how it affects distinguishing PD from H in our model at 2 seconds STFT window length and 20 output points. Transition from 1-20 Hz to 1-12 Hz and 1-8 Hz just slightly decreased the model performance, and it mostly affected 6 exercise (its recall metric dropped down to 0.85), while performance degradation for 3-8 Hz range in comparison with 1-8 Hz one was clear observed. Averaged metrics values within exercises dropped by 0.02-0.03 points, while recall value for exercise 6 dropped down to 0.8 (initially, it was 0.92 for 1-20 Hz range).

The final experiment was a comparison between the typical way of IMU data axes fusion, described in equation (1), and the proposed method (2). According to the previous experiments, 1-20 Hz frequency range was chosen to get the best results with a combination of 2 seconds STFT window length and 20 output points. The comparison results are presented in Table III in a format of mean ± 95% confidence interval (CI). Model performance was tested for all exercises. If the metric value of the proposed method is better, it is bold. If metrics of both methods are equal, they are both bold. Bottom row is an average value of each metrics over all exercises without CI, because CI is a pointless value here. As it can be observed, the proposed method (2) gives better result in 70% of cases, significantly outperforming the typical way (1) in some exercises.

IV. Conclusions

In this work, dominant hand invariant 1-D convolution model was proposed for distinguishing between PD and H and tested with different STFT window length, frequency ranges and number of output points, to find the optimal values of them and compare a novel STFT-based IMU data axes fusion method with a typical way. Results confirmed superiority of the proposed method in 70% of cases as well as efficiency of the proposed model with the best result of 98% of AUC and 92% of F1 and Accuracy metrics in 6 exercise. Poor accuracy of the rest tremor exercise 3, which is one of the cardinal PD symptoms, looks strange, however, during experiments, rest tremor amplitude was almost not observable in the most cases unlike the postural tremor in exercise 6, which can explain this performance contradiction. Remarkable thing is a good performance of 1 and 8 exercises, which are variable in details from patient to patient. Good performance here may justify the proposed idea of passing the entire exercise through the ML model, rather than slice it. Not-working condition of 7 exercise can be explained by prevalence of nut twisting movements over the tremor and spectrum overlapping, leading to a bad signal-to-noise ratio here. Good performance of 2,4 and 5 exercises is not surprising, but poor performance of 10 and 11 exercises was unexpected, because they are similar to 6 exercises (postural tremor). Perhaps, in depth, 6 exercise of outstretched arms position is indeed a special case, as it was mentioned in [10].

References

[1] V. L. Feigin, A. Abajobir, K. Abate, F. Abd-Allah, A. Abdulle, S. Abera, G. Abyu, M. Ahmed, A. Aichour, I. Aichour, *et al.*, "Gbd 2015 neurological disorders collaborator group. global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the global burden of disease study 2015," *Lancet Neurol*, vol. 16, no. 11, pp. 877–97, 2017.

[2] E. Dorsey, A. Elbaz, E. Nichols, F. Abd-Allah, A. Abdelalim, J. Adsuar, M. Ansha, C. Brayne, J. Choi, D. Collado-Mateo, *et al.*, "Gbd 2016 parkinson's disease collaborators. global, regional, and national burden of parkinson's disease, 1990–2016: a systematic analysis for the global burden of disease study 2016," *Lancet Neurol*, vol. 17, no. 11, pp. 939–953, 2018.

[3] W. A. Rocca, "The burden of parkinson's disease: a worldwide perspective," *The Lancet Neurology*, vol. 17, no. 11, pp. 928–929, 2018.

[4] J.-P. Bach, U. Ziegler, G. Deuschl, R. Dodel, and G. Doblhammer-Reiter, "Projected numbers of people with movement disorders in the years 2030 and 2050," *Movement disorders*, vol. 26, no. 12, pp. 2286–2290, 2011.

[5] A. Talitckii, E. Kovalenko, A. Shcherbak, A. Anikina, E. Bril, O. Zimniakova, M. Semenov, D. V. Dylov, and A. Somov, "Comparative study of wearable sensors, video, and handwriting to detect parkinson's disease," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.

[6] C. R. Baumann, "Epidemiology, diagnosis and differential diagnosis in parkinson's disease tremor," *Parkinsonism & related disorders*, vol. 18, pp. S90–S92, 2012.

[7] R. B. Postuma, D. Berg, M. Stern, W. Poewe, C. W. Olanow, W. Oertel, J. Obeso, K. Marek, I. Litvan, A. E. Lang, *et al.*, "Mds clinical diagnostic criteria for parkinson's disease," *Movement disorders*, vol. 30, no. 12, pp. 1591–1601, 2015.

[8] M. A. Thenganatt and E. D. Louis, "Distinguishing essential tremor from parkinson's disease: bedside tests and laboratory evaluations," *Expert review of neurotherapeutics*, vol. 12, no. 6, pp. 687–696, 2012.

[9] A. Gironell, B. Pascual-Sedano, I. Aracil, J. Marín-Lahoz, J. Pagonabarraga, J. Kulisevsky, *et al.*, "Tremor types in parkinson disease: a descriptive study using a new classification," *Parkinson's Disease*, vol. 2018, 2018.

[10] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of neurology, neurosurgery & psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[11] A. Talitckii, E. Kovalenko, A. Anikina, O. Zimniakova, M. Semenov, E. Bril, A. Shcherbak, D. V. Dylov, and A. Somov, "Avoiding misdiagnosis of parkinson's disease with the use of wearable sensors and artificial intelligence," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3738–3747, 2021.

[12] M. San Luciano, C. Wang, R. A. Ortega, Q. Yu, S. Boschung, J. Soto-Valencia, S. B. Bressman, R. B. Lipton, S. Pullman, and R. Saunders-Pullman, "Digitized spiral drawing: A possible biomarker for early parkinson's disease," *PloS one*, vol. 11, no. 10, p. e0162799, 2016.

[13] K. G. Sibley, C. Girges, E. Hoque, and T. Foltynie, "Video-based analyses of parkinson's disease severity: A brief review," *Journal of Parkinson's disease*, vol. 11, no. s1, pp. S83–S93, 2021.

[14] L. Giancardo, A. Sanchez-Ferro, T. Arroyo-Gallego, I. Butterworth, C. S. Mendoza, P. Montero, M. Matarazzo, J. A. Obeso, M. L. Gray, and R. Estépar, "Computer keyboard interaction as an indicator of early parkinson's disease," *Scientific reports*, vol. 6, no. 1, pp. 1–10, 2016.

[15] H. Gunduz, "Deep learning-based parkinson's disease classification using vocal feature sets," *IEEE Access*, vol. 7, pp. 115540–115551, 2019.

[16] H. Dai, G. Cai, Z. Lin, Z. Wang, and Q. Ye, "Validation of inertial sensing-based wearable device for tremor and bradykinesia quantification," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 4, pp. 997–1005, 2020.

[17] L. Tong, J. He, and L. Peng, "Cnn-based pd hand tremor detection using inertial sensors," *IEEE Sensors Letters*, vol. 5, no. 7, pp. 1–4, 2021.

[18] E. Samadi, H. Ahmadi, and F. N. Rahatabad, "Analysis of hand tremor in parkinson's disease: Frequency domain approach," *Frontiers in Biomedical Technologies*, vol. 7, no. 2, pp. 105–111, 2020.

[19] R. Deb, G. Bhat, S. An, H. Shill, and U. Y. Ogras, "Trends in technology usage for parkinson's disease assessment: A systematic review," *MedRxiv*, pp. 2021–02, 2021.

[20] C.-H. Lin, F.-C. Wang, T.-Y. Kuo, P.-W. Huang, S.-F. Chen, and L.-C. Fu, "Early detection of parkinson's disease by neural network models," *IEEE Access*, vol. 10, pp. 19033–19044, 2022.

[21] M. D. Nguyen, K.-R. Mun, D. Jung, J. Han, M. Park, J. Kim, and J. Kim, "Imu-based spectrogram approach with deep convolutional neural networks for gait classification," in *2020 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 1–6, IEEE, 2020.

[22] M. Barrachina-Fernández, A. M. Maitín, C. Sánchez-Ávila, and J. P. Romero, "Wearable technology to detect motor fluctuations in parkinson's disease patients: Current state and challenges," *Sensors*, vol. 21, no. 12, p. 4188, 2021.

[23] Y.-W. Kim, K.-L. Joa, H.-Y. Jeong, and S. Lee, "Wearable imu-based human activity recognition algorithm for clinical balance assessment using 1d-cnn and gru ensemble model," *Sensors*, vol. 21, no. 22, p. 7628, 2021.

[24] *SensorTile development kit*, STMicroelectronics.

[25] R. J. Elble and W. Ondo, "Tremor rating scales and laboratory tools for assessing tremor," *Journal of the Neurological Sciences*, p. 120202, 2022.

[26] Y. Zhou, M. E. Jenkins, M. D. Naish, and A. L. Trejos, "The measurement and analysis of parkinsonian hand tremor," in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 414–417, IEEE, 2016.

[27] R. J. Elble, "Estimating change in tremor amplitude using clinical ratings: recommendations for clinical trials," *Tremor and Other Hyperkinetic Movements*, vol. 8, 2018.