Objective Assessment of the Finger Tapping Task in Parkinson's Disease and Control Subjects using Azure Kinect and Machine Learning

Objective Assessment of the Finger Tapping Task in Parkinson's Disease and Control Subjects using Azure Kinect and Machine Learning

1st Gianluca Amprimo
Politecnico di Torino
Turin, Italy
gianluca.amprimo@polito.it

2nd Irene Rechichi
Politecnico di Torino
Turin, Italy
irene.rechichi@polito.it

3rd Claudia Ferraris
CNR-IEIIT Italy
Turin, Italy
claudia.ferraris@ieiit.cnr.it

4th Gabriella Olmo
Politecnico di Torino
Turin, Italy
gabriella.olmo@polito.it

***Abstract*—Parkinson's disease (PD) is characterised by a progressive worsening of motor functionalities. In particular, limited hand dexterity strongly correlates with PD diagnosis and staging. Objective detection of alterations in hand motor skills would allow, for example, prompt identification of the disease, its symptoms and the definition of adequate medical treatments. Among the clinical assessment tasks to diagnose and stage PD from hand impairment, the Finger Tapping (FT) task is a well-established tool. This preliminary study exploits a single RGB-Depth camera (Azure Kinect) and Google MediaPipe Hands to track and assess the Finger Tapping task. The system includes several stages. First, hand movements are tracked from FT video recordings and used to extract a series of clinically-relevant features. Then, the most significant features are selected and used to train and test several Machine Learning (ML) models, to distinguish subjects with PD from healthy controls. To test the proposed system, 35 PD subjects and 60 healthy volunteers were recruited. The best-performing ML model achieved a 94.4% Accuracy and 98.4% F1 score in a Leave-One-Subject-Out validation. Moreover, different clusters with respect to spatial and temporal variability in the FT trials among PD subjects were identified. This result suggests the possibility of exploiting the proposed system to perform an even finer identification of subgroups among the PD population.**

***Index Terms*—Parkinson's Disease, Finger Tapping, Pervasive Health, Telemedicine, Machine Learning, Azure Kinect, Mediapipe**

I. Introduction

Parkinson's Disease (PD) is a neurodegenerative disorder with a prevalence of 1% in the population over 65 years old [1], which is expected to rapidly increase as the global population grows older [2]. Patients affected by PD manifest alterations in motor functionalities – e.g., bradykinesia, akinesia, muscle stiffness, tremors, balance impairments – that worsen along with disease progression [3]. PD diagnosis and staging mainly relies on the assessment of motor symptoms during outpatient visits in clinical facilities, through well-standardised motor tasks, as those included in the Unified Parkinson's Disease Rating Scale (UPDRS) [4]. However, this kind of evaluation suffers from two main issues: on the one hand, different clinicians could disagree on the scoring assigned to the various tasks [5], since the evaluation is solely based on subjective observation and personal expertise. On the other hand, outpatient visits are infrequent (generally scheduled on a 6-month to 1-year basis); thus, the occurrence of specific symptoms might be identified and treated by clinicians only after several months from its onset. This clearly translates in a reduced quality of life for patients and decreased efficacy of the (therefore delayed) treatment. For these reasons, technological solutions for the objective diagnosis of the disease and the continuous follow up of PD patients, in a pervasive health scenario, are being largely investigated [6]–[8].

Among the standardised tasks of UPDRS, those related to hand impairment have a significant correlation with PD symptoms such as bradykinesia (i.e., reduced movement speed) and hypokinesia (i.e., reduced movement amplitude). Specifically, the Finger Tapping (FT) task is one of the most widely used by clinicians. It consists in tapping the tip of the thumb and index fingers as rapidly as possible, with the largest amplitude, for a fixed number of repetitions or seconds. Several studies aimed at objectively evaluating FT. Three main groups can be identified: (i) solutions based on wearable devices, such as inertial measurement units or instrumented gloves [9]–[11]; (ii) smartphone-based assessment [12]–[14], in which the interaction with the screen of the device is translated to an FT-equivalent; (iii) vision-based systems, using either RGB, RGB-Depth or Depth videocameras [15]–[17]. The first type of solutions are generally more invasive, might be less practical in unsupervised settings and therefore complex to be managed by patients on their own. Smartphone-based solutions only find indirect correspondence between the metrics collected by the touchscreen and the severity scores, and could be complex to interpret with respect to the standardised task. Finally, vision-based approaches exploit videocameras and markerless-tracking systems based on shallow or deep learning models, which first estimate a 2D or 3D hand skeleton and then evaluate a series of features describing FT. These features can be used to either classify subjects or estimate a severity score. The 2D tracking systems [18], [19] are easier to implement thanks to deep learning, but may provide limited insight into the real movement, since one relevant dimension is missing. At the same time, many 3D solutions still rely on some wearable component (e.g., gloves with markers [19], [20]) or allow for

tracking only inside a limited volume (e.g., using Leap sensor [21], [22]). Nevertheless, some recent works have proposed innovative 3D tracking systems which provide high level of accuracy, while leveraging simple Depth or RGB-Depth video streams [15]–[17].

This work investigated a possible system to objectively characterise FT from RGB-Depth video recordings, with the goal of automatically recognising PD and its symptoms (e.g., bradykinesia, hypokinesia). The system exploits an innovative approach for 3D-hand tracking (i.e., the GMH-D algorithm [15]) and Machine Learning (ML) methods to address this research problem. Moreover, an experimental session involving 60 healthy controls (HC) and 35 PD patients to test the system was performed, and it is reported in the remainder of the paper, along with its implementation and experimental results.

II. Materials and Methods

A. Acquisition & Processing System Pipeline

The overall system pipeline is described in Figure 1. The FT acquisition block is based on the GMH-D algorithm [15] for tracking hand joints during clinical assessment tasks involving hands. The fusion of the depth estimation provided by the Azure Kinect camera and the marker-less tracking provided by Google Mediapipe Hands [23] allows for the precise and objective tracking of hand joints trajectories also during highly dynamic tasks as FT. This property ensures high stability and precision in the estimation of features related to the motion of fingers. Moreover, the authors proved that the algorithm can be used to extract features to characterise FT at different speeds and with altered amplitude, achieving good automatic classification results in simulated trials performed by a group of healthy subjects.

Frontal recordings of FT tasks, lasting 10-seconds each, are acquired through a custom implementation of GMH-D developed in Unity® (Unity Technologies, San Francisco, CA, USA) running on a minipc equipped with a 9th generation Intel® CoreTM processor (2.4 GHz quad-core), 16 GB RAM, NVIDIA GeForce RTX 2060 6GB GDDR6, HDMI and USB3 ports, Windows 10 operating system. Video recordings are processed in real-time (30 frame per seconds) by the acquisition software, to produce in output a JSON-file containing the trajectories of 21 virtual hand joints. In the second block, offline data processing is carried out to: (i) compute the evolution over time of the distance between Index-Finger-Tip (IFT) joint and Thumb-Tip (TT) joint and then segment it, such that single FT movements are identified; (ii) extract a series of features describing the mean FT movements characteristics and their regularity over the whole task, through the coefficient of variation and spectral properties; (iii) select the optimal feature set (F_{opt}) to distinguish the PD subjects from the HC group. In the last block of the system, a classification based on optimal features is carried out through well-established shallow learning models. The comparison and evaluation of such models is performed by employing the Leave-One-Subject-Out (LOSO) procedure, to reduce the possible impact of model overfitting during the testing stage.

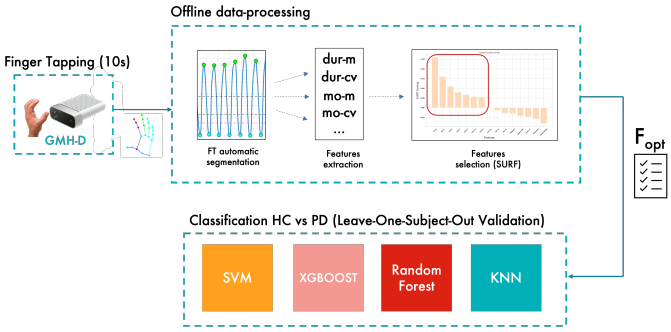

Fig. 1. Overall pipeline of the system proposed in this work.

B. Study Participants and Experimental Session

An experimental study was carried out to test the capability of the proposed system to automatically assess in an objective manner FT and its capability to support PD diagnosis and follow up. The experimental phase involved a total of 95 participants, including both people with PD and HC. A total of 35 PD subjects (15 females, 65.1 $\pm$ 9.2 years), enrolled at the association *"Associazione Amici Parkinsoniani Piemonte ONLUS"* (Turin, Italy) took part in the data collection. Both the recruitment and the experimental sessions occurred at the Association's Offices; hence, it was not possible to retrieve clinical information regarding the disease stage or the symptoms progression. The HC group, enrolled among the PD subjects' caregivers and the association personnel, comprised 60 people (27 females, 53.8 $\pm$ 7.9 years), with no history of neurological and cognitive disorders. Exclusion criteria, for both groups, included dementia or any psychiatric conditions that would prevent appropriate task completion.

The procedure has been conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of A.O.U. Città della Salute e della Scienza di Torino (Approval No. 00384/2020). All participants received detailed information on the study purpose and execution, and provided written informed consent for observational study.

As such, the experimental session was part of a broader observational study, entailing a high number of items and tasks in order to assess symptoms from different perspectives. Therefore, one single acquisition was performed by each subject involved in the study. The control group only performed the task with their dominant hand; on the contrary, as Parkinson's may lead to different impairment degrees in the left and the right upper limbs, the PD subjects underwent the task twice – i.e., with both the right and left hands, respectively. Moreover, since clinical scoring separately assesses the two limbs extremity, the left and right FT executions were considered as single data points in the subsequent analysis.

C. Data Pre-processing and Feature Extraction

All data were pre-processed prior to the feature extraction; both pre-processing and feature extraction were carried out with custom-written MATLAB® code (R2021b). Pre-processing was carried out on the 3D joint trajectories, and it

TABLE I
EXTRACTED KINEMATIC PARAMETERS FOR FT, ALONG WITH DESCRIPTION AND UNIT OF MEASURE.

Parameter	Description	Unit
mo_m	Range of motion as max - min amplitude (Mean)	mm
mo_{cv}	Range of motion as max - min amplitude (Variation Coefficient)	mm
ma_m	Maximum amplitude of movement (Mean)	mm
ma_{cv}	Maximum amplitude of movement (Variation Coefficient)	mm
mos_m	Maximum opening speed (Mean)	mm/s
mos_{cv}	Maximum opening speed (Variation Coefficient)	mm/s
mcs_m	Maximum closing speed (Mean)	mm/s
mcs_{cv}	Maximum closing speed (Variation Coefficient)	mm/s
dur_m	Movement duration (Mean)	s
dur_{cv}	Movement duration (Variation Coefficient)	s
freq_{low}	Frequency of voluntary movement	Hz
pow_{low}	Power of voluntary movement	$-$
freq_{band}	Frequency of tremor movement	Hz
pow_{band}	Power of tremor movement	$-$
pace_{var}	Variation in pace for low frequencies	$-$

involved resampling and filtering procedures. All signals were resampled at 50 fps to discard framerate jitter and increase the smoothness of the signal. A low-pass, zero-phase Butterworth order 3 filter (10 Hz cut-off frequency) was applied to decrease high-frequency noise; indeed, only the relevant human body-movements frequency band was retained in the analysis.

After pre-processing, the distance between IFT and TT joints was computed for the whole FT task. The obtained signal, which likely approximates a periodic sinusoidal signal in a healthy condition, is automatically segmented to identify single FT movements. The segmentation is carried out by means of a *maxima* and *minima* search on the signal, through proper thresholding. The points are then mapped to START, CLOSURE (during finger tapping) and END instants of single movements. Features are then extracted from both the single segments and the whole IFT-TT distance signal, following [20]. All extracted features are detailed in Table I. These parameters allow for the evaluation of hand impairment in PD and account for the span of the movements, in terms of time and amplitude. In more detail, they assess the duration of the FT exercise – from start to completion – the opening and closing velocity, as well as the frequency of the voluntary movements and tremor, and any variation in pace. In particular, the distinction between closing and opening speed is a finer description that is not commonly considered, since usually only the overall velocity is evaluated. As regards the amplitude, the extracted parameters describe the movements range and the opening amplitude. Given that the extracted parameters are described on different magnitude scales, z-score normalisation was applied to all features to allow for proper implementation of the ML algorithms.

D. Feature inspection, selection and automatic classification

The computed features were first inspected through statistical testing to evaluate their distribution in the two groups. This analysis was conducted using the open-source statistical tool Jamovi [24]. First, the normality of the features was investigated by means of the Shapiro-Wilk test. Since all features resulted to be non-normally distributed, a non-parametric approach was selected, employing the Mann-Whitney independent sample U-Test, to identify characteristics differently distributed between the PD and HC groups. Finally, Spearman's correlation was also computed to evaluate the degree of linear dependency between the exploited features and the group label (either PD or HC).

Feature selection was performed by means of the the Speeded Up Robust Features (SURF) algorithm [25], by selecting only those features with a positive score. This reduced feature set F_{opt} was employed for classification.

To automatically classify FT executions by PD and HC subjects, four supervised models were explored and implemented in Python, namely: Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Random Forest (RF) and Extreme Gradient Boosting (XGBoost). Since the overall dataset contains few subjects (< 100) and one or two FT trials (at most) associated to each participant in the experimental session, a LOSO cross-validation procedure was implemented, as previously mentioned. In LOSO, for each investigated shallow method, $N-1$ models are trained, where N is the number of subjects involved in the study. Hence, each model is trained on the whole dataset, except for the holdout subject – which is then used for testing. Model scoring metrics, namely Accuracy, Precision, Recall and F1-score, are evaluated on the results obtained by all models. The optimal hyperparameters for the classification models were identified using k-fold Cross Validation (CV) (k = 3) combined with a Grid search approach on the training data only, to further ensure the robustness of the trained classifiers. A complete summary of the employed models and the optimised parameters is provided in Table II; the table displays, for each model, the hyperparameters search range. Finally, considering the clinical relevance of the spatial and temporal variability in the execution of the FT movements – described by the features mo_{cv} and dur_{cv} – the PD group was further investigated along these two variables. Indeed, they describe the extent of motor control in the execution of a repetitive and fine movement such as the FT. The variables mo_{cv} and dur_{cv} were taken as axes, and clusterisation was applied to the data in the PD group, through Agglomerative Clustering. This may be considered a preliminary step to prove the feasibility of the system of recognising subtypes of PD symptoms related to these two properties (i.e., bradykinesia, hypokinesia).

III. RESULTS

A. Statistical analysis and features selection

Table III shows the results for the statistical analysis of the extracted features (normality testing by Shapiro-Wilk and distribution differences in the two classes by Mann-Whitney U Test), as well as their correlation with the output label – i.e., the presence of PD. First of all, as it can be appreciated from the numerical results, all features are not normally distributed

TABLE II
SUMMARY OF THE EMPLOYED CLASSIFIERS, THE SEARCHED HYPERPARAMETERS (PARAMETER AND SEARCH RANGE) AND THE OPTIMISED CONFIGURATION IN THE FINAL MODEL.

Model	Searched Hyperparameters	Optimised
SVM	*Kernel function*: linear, polynomial, radial basis, sigmoid *Penalty (C)*: [0.1, 1, 10, 100, 1000] γ: [1, 0.1, 0.001, 0.0001]	Radial basis function C = 1 γ = 1
KNN	*Minkowski Distance order (p)*: [1, 2, 3, 4, 5] *Number of neighbours (K)*: [3, 5, 7] *Weights (W)*: uniform, distance-based	p = 2 K = 5 W: distance-based
RF	*Number of trees*: [50, 100, 150] *Depth*: [3, 5, 7] *Minimum Features (leaf)*: [1, 2, 4] *Minimum Features (node)*: [2, 5, 10]	N_{trees} = 100 Depth = 3 N_{leaf} = 1 N_{split} = 2
XGBoost	*Number of trees*: [50, 100, 150] *Depth*: [3, 5, 7] *Learning rate*: [0.001, 0.05, 0.1]	N_{trees} = 100 Depth = 3 Learning rate = 0.1

– indeed, small values of p for the Shapiro-Wilk Test suggest violation of the normality hypothesis. In the Mann-Whitney U Test, the p-values suggest that all features are discriminating well between PD and HC, with the exception of mcs_m, mcs_{cv} and freq_{band}, which also show a fairly low Spearman's ρ. The first two parameters refer to the closing velocity during tapping. Overall, PD subjects are expected not to be challenged by the closing part of the movement, but rather in the starting phase (opening of the thumb and index), as the disease is often characterised by difficulty in starting motion with body limbs. Indeed, the equivalent parameters for the opening velocity (mos_m and mos_{cv}) show statistical significance in the test; hence, they are differently distributed in the PD and HC groups. Regarding freq_{band}, this feature is related to the frequency components of tremor. Nevertheless, PD is mainly identified by resting tremors, which tends to disappear during the execution of specific movements. In addition to this, few of the recruited subjects showed this very specific symptom.

Figure 2 shows the ranking of the features, as provided by the SURF algorithm. For the F_{opt}, only the features with a positive score were selected, leading to a final set of size 7. All selected features were significant in the statistical analysis and with a high-to-moderate correlation to the classification label. For the sake of completeness, Figure 3 reports the normalised violin plots estimated using the Kernel Density Estimation (KDE) approach, which further confirm the difference in the distribution of the optimal features between HC and PD.

B. Classification: LOSO Performance

Table IV reports the LOSO performance yielded by the employed classifiers, in terms of Accuracy, Recall, Precision and F1 score. All classifiers present with overall Accuracy values over 90%, which are indicative of very good classification performance. Accordingly, the F1 score – computed as the harmonic mean of Recall and Precision – is over 90 %.

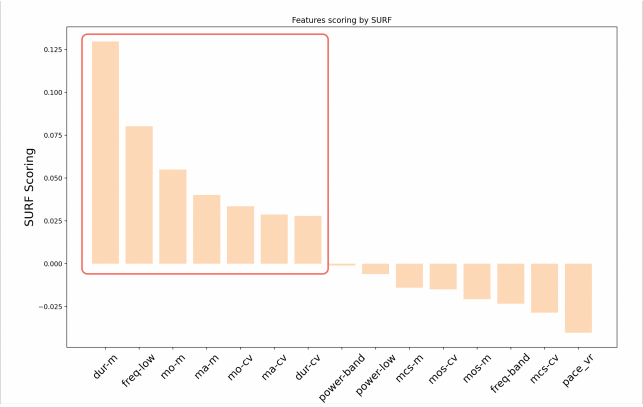

Fig. 2. Features rank scores yielded by the SURF algorithm. The selected features (positive scores) are squared in red.

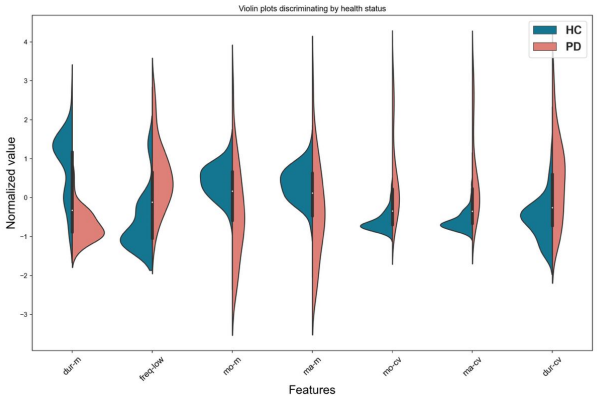

Fig. 3. Normalised violin plots of features in F_{opt}, estimated using the KDE approach, for PD and HC.

This, in particular, implies that the models are robust against both false positives and false negatives, suggesting that the employed features are highly predictive of PD. The overall

TABLE III
STATISTICS OF THE FEATURES EMPLOYED IN THE CLASSIFICATION. LEVEL OF SIGNIFICANCE: ***: p-VALUE < 0.001; **: p-VALUE < 0.01; *: p-VALUE < 0.05.

Feature	Shapiro-Wilk W	Mann-Whitney U	Spearman's ρ
mo_m	0.965**	1001***	-0.419*
mo_{cv}	0.846***	313***	0.725*
ma_m	0.963**	1131***	-0.361*
ma_{cv}	0.826***	316***	0.723*
mos_m	0.976**	1378**	0.251**
mos_{cv}	0.809***	1448*	0.220*
mcs_m	0.965**	1553	0.173
mcs_{cv}	0.807***	1917	0.012
dur_m	0.961***	265***	-0.746***
dur_{cv}	0.974*	830***	0.495***
freq_{low}	0.913***	543***	0.625***
power_{low}	0.518***	476***	-0.652***
freq_{band}	0.957***	1542	0.170
power_{band}	0.380***	1142***	0.356***
pace_{var}	0.923***	1293**	-0.291***

TABLE IV
LOSO CLASSIFICATION PERFORMANCE OF THE EMPLOYED CLASSIFIERS.

	SVM	RF	KNN	XGBoost
Accuracy	92.8 %	90.4 %	94.4 %	91.2 %
Recall	94.0 %	89.6 %	91.0 %	89.6 %
Precision	92.6 %	92.3 %	98.4 %	93.8 %
F-1	93.3 %	90.9 %	94.6 %	91.6 %

best performing model is KNN (K=3, p=2 – i.e., Euclidean Distance). This result is likely related to the fact that the SURF algorithm, employed in the feature selection stage, is based on the Nearest Neighbour concept; hence, it can be assumed that the obtained F_{opt} is remarkably accurate in combination with such model.

C. PD group inspection by clustering

Starting from the highly positive results that were achieved in the binary classification, an unsupervised approach was selected to further investigate the FT trials in PD subjects. In particular, the analysis focused on studying the variability in the excursion (mo_{cv}) and duration of the single movements (dur_{cv}), for two main reasons.

First, they represent the two axes that are implicitly evaluated by the clinicians during outpatient visits (regularity in the spatial and temporal property of motion). Second, these two features were found relevant both in the statistical analysis and in the feature selection procedures, respectively. In particular, the duration was selected as it accounts both for the variation in velocity (faster movement, shorter duration and viceversa) and in frequency properties. Figure 4 shows a scatter plot of the two variables, using as colourmap for the data points the clusterisation labels yielded by the Agglomerative Clustering – carried out on the two selected features only. The optimal number of clusters was set to 3 after the inspection of the dendrogram built by the clustering algorithm.

As it can be appreciated, three main subgroups were identified:

- **C0**: this cluster corresponds to low spatial variability in the execution of the movement, but medium to high duration variability, therefore reduced regularity in the speed of tapping. This group is likely to include subjects affected by bradykinesia – i.e., slowness of movement and speed (or progressive hesitations/halts) as movements are continued;
- **C1**: this cluster shows high spatial variability and medium duration variability. Subjects included in this cluster are likely to be either affected by hypokinesia – i.e., the movements amplitude is limited, especially in repeated sequences – and/or subjects with reduced control in the opening phase;
- **C2**: this cluster shows reduced both spatial and temporal variability; thus, movements are characterised by a good level of regularity. This implies that the group includes subjects with a lower level of hand impairment.

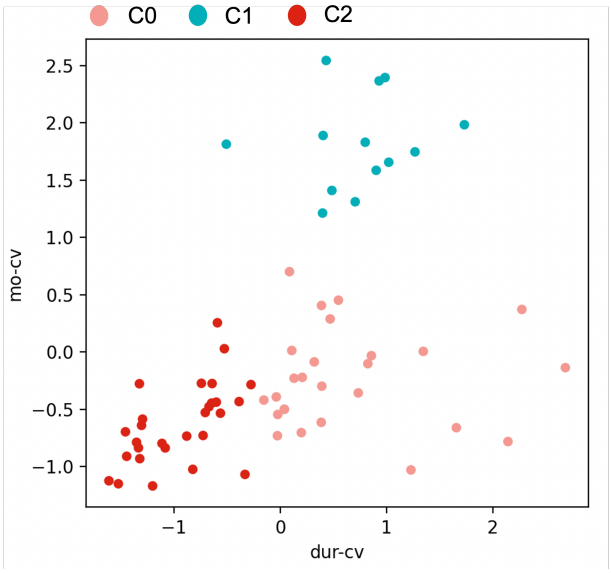

Fig. 4. Scatter plot of FT trials by PD subjects according to (mo_{cv}) and (dur_{cv}). Colours refer to clusters identified using the Agglomerative clustering approach.

Overall, these results seem to suggest that the proposed system and these two specific features evaluated by it could allow for a finer classification of the PD subjects into specific subgroups, exhibiting different symptoms related to hand and overall body movements.

IV. CONCLUSIONS

Parkinson's Disease is a progressive neurodegenerative disorder whose diagnosis is solely based on symptomatic and subjective evaluation carried out by physicians. The ideal follow-up procedures for the disease would require a continuous monitoring approach, whereas actual outpatient protocols are infrequent and might lead to late identification of new symptoms' onset.

Among the motor tasks that are generally evaluated in clinical settings, Finger Tapping is one of the most common, easy to perform and assess. Besides, it is suitable for smooth implementation in a pervasive health scenario, where video recordings of the task can be easily collected and processed.

This works proposes a non-invasive and easy-to-implement pipeline based on the GMH-D algorithm, to objectively characterise Finger Tapping from video recordings. This is done in order to extract a series of relevant features that could be used to automatically identify PD and some specific symptoms, such as bradykinesia or hypokinesia. The system estimates a series of spatial, temporal and frequency features of the evolution of the distance between the thumb and the index fingers during an FT trial lasting 10 seconds.

From the statistical analysis, the features proved to be differently distributed across the PD and HC subjects that were recruited to test the system, with a high correlation with the PD status (Section III-A). A robust feature selection method (SURF algorithm) was employed to select the optimal

features set F_{opt}, which included features related to single FT movements' spatial amplitude, duration, and frequency of voluntary movements. Four Machine Learning models, namely SVM, XGBoost, RF and KNN, were trained on this subset; they all achieved Accuracy, F1 score, Precision and Recall values above 90% in a Leave-One-Subject-Out validation. The work most suitable for a direct comparison [16], which similarly exploited a completely markerless 3D solution for tracking FT, reports a 76.9% Accuracy in a LOSO validation – though this value refers to a 4 severity-groups classification task. The results attained through the system proposed in this work seem to be in line with [16], taking into account the fact that this paper performs a binary classification only (PD vs HC). As a limitation, in this work the UPDRS scores of the single trials were not available. Future work will focus on investigating the possibility of regressing the UPDRS score, in accordance with the clinicians' evaluation, through the system pipeline proposed in this paper, as similarly performed in [16]. Nevertheless, the preliminary exploration of the PD group through unsupervised learning (Agglomerative clustering, Section III-C) already suggested that different subgroups in the PD population may be detected with respect to the irregularity in the spatial amplitude and duration of the single FT movements. The identified clusters seem to describe different symptoms occurring in the patients, such as bradykinesia and hypokinesia. This finer analysis represents a novelty with respect to [16], that did not investigate specific alterations occurring in the PD population. The recognition of these subtypes of PD represents an interesting and challenging research direction to be further investigated in future works.

The results presented in this paper, though preliminary, suggest that the proposed system may be a useful tool to support clinicians in the assessment of symptoms severity, thus bolstering the implementation of objective methods to evaluate not only FT but also PD itself, and paving the way to pervasive-health-based follow-up procedures.

REFERENCES

[1] D. K. Simon, C. M. Tanner, and P. Brundin, "Parkinson disease epidemiology, pathology, genetics, and pathophysiology," *Clinics in geriatric medicine*, vol. 36, no. 1, pp. 1–12.

[2] A. Association *et al.*, "2018 alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 14, no. 3, pp. 367–429, 2018.

[3] S. Sveinbjornsdottir, "The clinical symptoms of parkinson's disease," *J. Neurochem.*, vol. 139, pp. 318–324, 2016.

[4] C. G. Goetz, B. C. Tilley, S. R. Shaftman, Stebbins, *et al.*, "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results: MDS-UPDRS: Clinimetric assessment," *Mov. Disord.*, vol. 23, no. 15, pp. 2129–2170, 2008.

[5] C. G. Goetz and G. T. Stebbins, "Assuring interrater reliability for the updrs motor section: utility of the updrs teaching tape," *Movement Disorders*, vol. 19, no. 12, pp. 1453–1456, 2004.

[6] C. Ferraris, I. Ronga, R. Pratola, G. Coppo, T. Bosso, S. Falco, G. Amprimo, G. Pettiti, S. Lo Priore, L. Priano, A. Mauro, and D. Desideri, "Usability of the REHOME solution for the telerehabilitation in neurological diseases: Preliminary results on motor and cognitive platforms," *Sensors (Basel)*, vol. 22, no. 23, p. 9467, 2022.

[7] C. Godinho, J. Domingos, G. Cunha, A. T. Santos, R. M. Fernandes, D. Abreu, N. Gonçalves, H. Matthews, T. Isaacs, J. Duffen, *et al.*, "A systematic review of the characteristics and validity of monitoring technologies to assess parkinson's disease," *Journal of neuroengineering and rehabilitation*, vol. 13, no. 1, pp. 1–10, 2016.

[8] E. Kuosmanen, F. Wolling, J. Vega, V. Kan, Y. Nishiyama, S. Harper, K. Van Laerhoven, S. Hosio, D. Ferreira, *et al.*, "Smartphone-based monitoring of parkinson disease: quasi-experimental study to quantify hand tremor severity and medication effectiveness," *JMIR mHealth and uHealth*, vol. 8, no. 11, p. e21543, 2020.

[9] L. di Biase, S. Summa, J. Tosi, F. Taffoni, M. Marano, A. Cascio Rizzo, F. Vecchio, D. Formica, V. Di Lazzaro, G. Di Pino, and M. Tombini, "Quantitative analysis of bradykinesia and rigidity in parkinson's disease," *Front. Neurol.*, vol. 9, 2018.

[10] F. Cavallo, A. Moschetti, D. Esposito, C. Maremmani, and E. Rovini, "Upper limb motor pre-clinical assessment in parkinson's disease using machine learning," *Parkinsonism Relat. Disord.*, vol. 63, pp. 111–116, 2019.

[11] V. Bobić, M. Djurić-Jovičić, N. Dragašević, M. B. Popović, V. S. Kostić, and G. Kvaščev, "An expert system for quantification of bradykinesia based on wearable inertial sensors," *Sensors (Basel)*, vol. 19, no. 11, p. 2644, 2019.

[12] L. Omberg, E. Chaibub Neto, T. M. Perumal, A. Pratap, A. Tediarjo, J. Adams, B. R. Bloem, B. M. Bot, M. Elson, S. M. Goldman, *et al.*, "Remote smartphone monitoring of parkinson's disease and individual response to therapy," *Nature Biotechnology*, vol. 40, no. 4, pp. 480–487, 2022.

[13] A. Zhan, S. Mohan, C. Tarolli, R. B. Schneider, J. L. Adams, S. Sharma, M. J. Elson, K. L. Spear, A. M. Glidden, M. A. Little, *et al.*, "Using smartphones and machine learning to quantify parkinson disease severity: the mobile parkinson disease score," *JAMA neurology*, vol. 75, no. 7, pp. 876–880, 2018.

[14] A. Lauraitis, R. Maskeliūnas, R. Damaševičius, and T. Krilavičius, "A mobile application for smart computer-aided self-administered testing of cognition, speech, and motor impairment," *Sensors*, vol. 20, no. 11, p. 3236, 2020.

[15] G. Amprimo, C. Ferraris, G. Masi, G. Pettiti, and L. Priano, "GMH-D: Combining google MediaPipe and RGB-depth cameras for hand motor skills remote assessment," in *2022 IEEE International Conference on Digital Health (ICDH)*, IEEE, 2022.

[16] Z. Guo, W. Zeng, T. Yu, Y. Xu, Y. Xiao, X. Cao, and Z. Cao, "Vision-based finger tapping test in patients with parkinson's disease via spatial-temporal 3d hand pose estimation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3848–3859, 2022.

[17] P. J. M. Bank, J. Marinus, C. G. M. Meskers, J. H. de Groot, and J. J. van Hilten, "Optical hand tracking: A novel technique for the assessment of bradykinesia in parkinson's disease," *Mov. Disord. Clin. Pract.*, vol. 4, no. 6, pp. 875–883, 2017.

[18] Y. Liu, J. Chen, C. Hu, Y. Ma, D. Ge, S. Miao, Y. Xue, and L. Li, "Vision-based method for automatic quantification of parkinsonian bradykinesia," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 10, pp. 1952–1961, 2019.

[19] S. Williams, Z. Zhao, A. Hafeez, D. C. Wong, S. D. Relton, H. Fang, and J. E. Alty, "The discerning eye of computer vision: Can it measure parkinson's finger tap bradykinesia?," *Journal of the Neurological Sciences*, vol. 416, p. 117003, 2020.

[20] C. Ferraris, R. Nerino, A. Chimienti, G. Pettiti, N. Cau, V. Cimolin, C. Azzaro, G. Albani, L. Priano, and A. Mauro, "A self-managed system for automated assessment of UPDRS upper limb tasks in parkinson's disease," *Sensors (Basel)*, vol. 18, no. 10, p. 3523, 2018.

[21] A. Moshkova, A. Samorodov, N. Voinova, A. Volkov, E. Ivanova, and E. Fedotova, "Parkinson's disease detection by using machine learning algorithms and hand movement signal from LeapMotion sensor," in *2020 26th Conference of Open Innovations Association (FRUCT)*, IEEE, 2020.

[22] A. H. Butt, E. Rovini, C. Dolciotti, G. De Petris, P. Bongioanni, M. C. Carboncini, and F. Cavallo, "Objective and automatic classification of parkinson disease with leap motion controller," *Biomed. Eng. Online*, vol. 17, no. 1, p. 168, 2018.

[23] "Mediapipe hands." https://google.github.io/mediapipe/solutions/hands.html. Accessed: 2022-12-14.

[24] "Jamovi - open statistical software for the desktop and cloud." https://www.jamovi.org. Accessed: 2022-12-19.

[25] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*, pp. 404–417, Springer, 2006.