

Towards a Comprehensive Solution for Vision-based Digital Neurological Examination

Trung-Hieu Hoang*, Mona Zehni*, Huaijin Xu, George Heintz, Christopher Zallek, Minh N. Do

Abstract—Objective: Teleneurology systems have gained significant adoption over the past decades due to their potential in expanding access to neurology care despite the unequal geographic distribution of neurology resources. Current teleneurology pipelines are narrowed down to a specific neurological exam subtype and are not designed for home application. In this paper, we propose an accessible vision-based tele-exam and documentation solution called Digital Neurological Examination (DNE) to address those limitations. **Method:** Through our DNE software, clinical personnel and people at home are enabled to video capture an examination while performing instructed neurological tests, including finger tapping, finger to finger, forearm roll and stand-up and walk. Our modular design of the DNE software supports integrations of additional examinations. The DNE extracts from the recorded examinations the 2D/3D pose and quantifies kinematic and spatio-temporal features. The features are clinically relevant and allow a physician to quantify movement impairments and changes over time. A web server and a user interface for recordings viewing, feature visualizations, and annotation is available. **Results:** The DNE was developed based on a collected data set of 21 subjects containing normal and simulated-impaired movements. The overall accuracy of the system is demonstrated by classifying the recorded movement using various machine learning models. Our tests show an accuracy beyond 90% for upper-limb tests and 88% for the stand and walk tests.

Index Terms—Teleneurology (TN), quantitative analysis, disease documentation, monitoring, finger tapping, finger to finger, forearm roll, stand and walk, gait, human pose, machine learning.

I. INTRODUCTION

The burden and prevalence of neurological disorders [1] and the national shortage of neurologists [2] continue to grow hand in hand. This increases disparity through unequal access to clinical care and drives worsening clinician burnout rates. Meanwhile, the COVID-19 pandemic has boosted the transition from in-person to virtual neurological examinations [3], [4]. Teleneurology (TN), providing virtual consultations and clinical care through video-conferencing systems or mobile

devices [5] to patients with neurological disorders, similar to in-person assessments [6], is developing rapidly and shows the potential to make efficient assessments remotely [7]–[9]. TN helps to distribute scarce healthcare resources and enhance accessibility to neurological care [10], [11]. In addition, TN solutions with quantification of physical evaluations that bypass clinician availability and subjectivity of assessments [12] are important to improve care and compensate for the shortage of neurologists.

Current teleneurology systems are devoted to a single neurological test [13]–[15], require advanced setups/equipment [16], or lack automated assessments [17]. Therefore, a teleneurology solution, 1) suitable for use by neurologists and non-neurologists, 2) with wide applicability at clinics or home, 3) that is easy to deploy, 4) supports a wide range of neurological tests, and 5) enables automated objective quantitative evaluations would significantly advance health care delivery.

For this purpose, in this work, we introduce an end-to-end vision-based TN platform named Digital Neurological Examination (DNE). As part of DNE, we designed an easy-to-use smartphone software with pre-defined examination instructions. The DNE software allows the users to video record their performance on several neurological screening examinations, including finger tapping (FT), finger to finger (FTF), forearm roll (FR), and stand and walk (SAW). These recordings are uploaded to secure cloud-based storage. In an offline step, for each recording, 2D/3D pose, estimating the location of major human body keypoints is extracted using off-the-shelf deep-learning-based solutions such as OpenPose [18] and VideoPose3D [19]. From the estimated pose, unified digital bio-markers, including spatio-temporal and kinematic features, are computed [20]. We showcase the performance of our system on a dataset collected from 21 healthy subjects taking different neurological screening tests (FT, FTF, FR, SAW) when their function is normal or with a simulated impairment. We incorporate our defined features in a variety of machine learning models to detect abnormal functioning in our dataset. Fig. 1 further illustrates our DNE system.

We summarize the key contributions of this work as: (1) We developed a unified and modular software package for high-quality DNE recording collection. Our DNE software is easy-to-use, allows the integration of new tests, and runs on handheld iOS devices. We also implemented a web-based dashboard for viewing the recordings and feature visualization. (2) We propose a vision-based approach to study various neurological tests, including FT, FTF, FR, and SAW, to simu-

* Both authors contributed equally.

Trung-Hieu Hoang, Mona Zehni and Minh N. Do are with ECE and CSL department at UIUC.

Huaijin Xu is with Department of Kinesiology and Community Health at UIUC.

George Heintz is with Healthcare Engineering Systems Center at UIUC. Christopher Zallek is with OSF HealthCare Illinois Neurological Institute - Neurology.

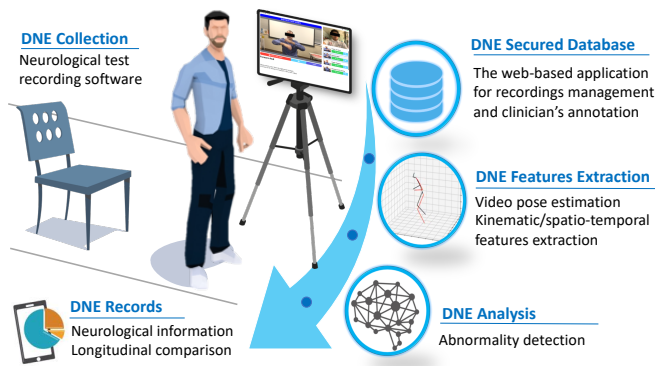


Fig. 1: Illustration of our digital neurological exam (DNE) system. Users video record their performance on various neurological tests. The recordings are uploaded to secured cloud storage as a DNE record. Pose data are extracted from the videos and used to calculate kinematic/spatio-temporal features. These features are used in abnormality detection and longitudinal analysis.

late whole-body physical assessments in clinical settings. For each test, we define separate kinematic and spatio-temporal quantified features. We believe these features advance clinicians' disease documentation while promoting self-monitoring and early symptom detection [21], [22]. Furthermore, collecting patients' recordings alongside the quantified features at different time points facilitates supervision and monitoring of disease progression over time [23]. (3) To the best of our knowledge, we are the first to construct a vision-based dataset consisting of multiple neurological tests and simulated-impaired video recordings per subject alongside the extracted 2D/3D pose. Analyzing this dataset allows us to have a normal self-baseline for each abnormal recording and test the power of the extracted features in distinguishing normal from abnormal performance. Our dataset (excluding RGB videos due to privacy restrictions) will be publicly available for future research.

The organization of this paper is as follows. Section II summarizes recent studies on teleneurology systems. Section III describes DNE's software platform used in our data collection. Section IV introduces our DNE dataset. We define our features in detail in Section V. Section VI contains our analysis results while Section VII draws our main conclusions.

II. RELATED WORK

In this section, we review the related literature to different tests including FT, FTF, FR and SAW. For each test, we briefly discuss the existing sensor, web/smartphone and vision based solutions.

Finger Tapping (FT): Sensor-based FT assessments study spectral analysis of gyroscope data [24], opening finger tap velocity captured by accelerometers [25], standard deviation, range and entropy measured by a collection of sensors including synchronized wrist watches, pressure sensors

and accelerometers [26]. Several smartphone based applications [27]–[30] are designed to quantitatively evaluate various symptoms and motor skills in patients with Parkinsons Disease (PD). While these approaches are proven effective and low cost, their measurements are not as informative as vision based methods, relying on video data and simulating in-person clinical examinations.

Among vision based pipelines, [13], [31]–[33] extract a set of kinematic interpretable features from the tracked positions of the fingers given an RGB video. These features are easy to explain and associate with clinical symptoms. On the other hand, black box deep learning models operating on the estimated finger poses and their derivatives are proposed in [34]. While these solutions provide high accuracy, unlike our DNE analysis pipeline, they lack explainability and require large training sets to generalize and avoid overfitting.

Finger to Finger (FTF): A well-studied test in the literature that is similar to FTF in terms of measuring smoothness and upper extremity coordination is the finger to nose test. Among sensor-based methods, Rodrigues et al in [35] investigates the coordination ability of patients with chronic stroke versus healthy control using a complex marker-based motion analysis system. Oubre et al [36] studied ataxia through wearable inertial sensors and a computer tablet version of finger to nose test. Furthermore, predicting severity levels of ataxia or PD via a rapid web-based computer mouse test is explored in [37]. Jaroensri et al. [14] is among the first to propose vision based solutions that are on par with a specialist in terms of rating the severity scale of PD while relying on estimated joint positions from recorded videos.

Upper Limb Tests: To the best of our knowledge, sensor-based or vision-based studies related to the forearm roll task are scarce. Thus, here we further overview the existing methods devoted to the study of upper limb movements. Using wearable sensors, Cruz et al in [16] assessed the acceleration, velocity or smoothness of the upper limb motor function of patients after stroke. A low-cost Kinect based solution, tracking subjects' hand when asked to move a marker on a rectangular pattern is proposed in [38]. The range of motion is analyzed using an internet-based goniometer in [39]. In [40], the authors describe a vision based system that captures upper limb motions via multiple cameras installed at different views. While this multi-view multi-camera system is less sensitive to occlusions and dynamic backgrounds, unlike our DNE system, it requires a special set-up which is hard to install for home-use.

Stand and Walk (SAW): In our review of gait analysis literature, we focus on the marker-less [41] vision-based solutions, mainly measured using general handheld cameras and mobile devices. In early efforts for marker-less gait analysis, silhouettes are extensively used to detect heel-strike and toe-off occurrences. These two events refer to the first and last ground contact of each foot, later on adopted to accurately estimate important gait parameters [42]–[45]. However, these methods are restricted to specific laboratory settings and are sensitive to the quality of foreground/background segmentation. The surge of research in the human pose estimation field [46]–[48] brought along popular deep learning frameworks which accu-

rately estimate the 2D/3D location of body joints from different inputs including RGB image, video and depth maps [18], [19], [49], [50]. Depth-map based gait assessment solutions relying on the estimated pose from either depth or RGBD [51], [52], have studied the rotational angle and angle velocity of certain body keypoints [53] and evaluated the spatio-temporal gait metrics such as step length and time [15], [54].

Using the estimated pose from OpenPose [18], Xue *et al.* [15] studied the remote monitoring of gait parameters for senior care. To assess the freezing of gait (FoG) symptom in Parkinson patients, [55] proposed the use of frequency analysis methods while [56] adopted graph convolutional neural networks to attain the probability of FoG from pose data. Kidziński in [57] employed black-box deep learning models to estimate the level of movement disorder in children suffering from cerebral palsy. Despite their promising results, deep-learning based solutions are less interpretable and require large training supervised datasets for better generalization.

III. SYSTEM DESIGN

As part of DNE, we developed three software packages described below, to maintain data acquisition, analysis and results report.

DNE Recorder: This module accommodates easy-to-use self or assisted video recording on a set of pre-defined neurological examinations. We implemented DNE Recorder as a mobile application, supported on handheld devices with iOS 13 and above.

This software includes detailed instructions on how to perform each test alongside automated video capturing functions. Our software facilitates recording of high quality depth maps on iOS devices equipped with LiDAR. We collect 1080×720 high-quality RGB and depth videos (upon applicable hardware) at 60 frames per second (FPS). We also record camera intrinsic and extrinsic parameters. All recordings are synchronized into a secure cloud storage for offline processing. The user interface of this module is shown in Fig. 2a.

DNE Analyzer: We analyze the RGB recordings offline in a separate module. The main components of DNE Analyzer include 1) off-the-shelf vision based pose estimation, 2) feature extraction, 3) abnormality detection. Section V is devoted to an elaborate description of this module.

DNE Viewer: Lastly, we provide a secure web application for clinicians, neurologists and researchers to monitor and analyze raw recordings from all subjects remotely. Furthermore, we provide a separate dashboard to view the analysis results for each DNE recording. Fig. 2b displays screenshot examples of the user interface of DNE Viewer.

IV. DATASET COLLECTION

Our dataset collection protocol is IRB approved (#IRB.1452500) [Hieu: add the approval date] by the University of Illinois College of Medicine at Peoria Institute Review Board 1. In this study, 21 healthy volunteers (18 females and 3 males) were recruited by sampling of convenience and examined at the OSF HealthCare Illinois Neurological

TABLE I: Summary of our DNE dataset

Test	Total	Label		View		Video	
		Normal	Abnormal	Front	Side	RGB/D	RGB
FT	68	32	36	68	-	32	36
FR	78	39	39	78	-	32	46
FTF	88	41	47	88	-	40	48
SAW	100	44	56	55	45	41	59

Institute Outpatient Neurology Clinic (Peoria, IL). Each subject's testing lasted for thirty to forty-five minutes.

Neurological examinations included in DNE test fine motor and mobility abilities. For fine motor tasks, we study Finger Tapping (FT), Finger to Finger (FTF) and Forearm Rolling (FR) tests. We evaluate mobility by the Stand and Walk (SAW) test. In the FT test, participants are instructed to put their arms and hands within the camera view when their index fingers and thumbs were touched. Then they would start tapping them as big open and close, and fast as they could for 15 seconds. In FTF, participants repetitively first point their index fingers towards the ceiling and then touch their fingers together out in front of their chests with their eyes open for a duration of 15 seconds. In the FR test, participants are asked to gently clench their hands, hold their forearms horizontally, and roll their hands around each other as fast as possible for 15 seconds. In the SAW test, participants stand from a sitting pose in a chair, move the chair out of the way, walk back and forth 15 feet. The designated time for SAW recording is 45 seconds.

Each subject took two sets of neurological examinations supervised by a neurologist. In the first set of examinations, the subjects performed the tasks normally. However, for the second set of recordings, the subjects were asked to simulate motor dysfunction, i.e. perform the test abnormally. For this purpose, the subjects wore devices to deliberately add disruption to their performance and mimic impairments. For FT, a rubber band is used to restrict movements of the index and thumb fingers. For the FR and SAW tests the subjects put on a left wrist and a knee brace, respectively. On the other hand, for the FTF test, the subjects were asked to deliberately mimic a tremor pattern in moving their fingers and hands. Snapshots of recordings and subjects wearing the devices are exhibited in Fig. 3.

Both set of recordings are acquired by our DNE Recorder on iPad 11 Pro and iPhone devices. For upper body tests, we have a close-up frontal view of the subjects with visible pelvis. Moreover, to assess the invariance of our analysis under small deviations from the frontal camera view, the view of the recordings taken on iPhone is slightly to the left compared to the iPad recordings. In addition, for the SAW test, we record both sagittal and frontal views, using iPad and iPhone, respectively. In total, including all four tests (FR, FT, FTF, SAW), we collect 334 videos. Table I provides a summary of our dataset.

While there is hardly any similar publicly available upper-body neurological related dataset, there are several datasets studying gait impairments specifically in [15], [42], [43], [55], [57]. The closest to our dataset is KIMORE [58] focusing on rehabilitation exercises rather than neurological tests. The

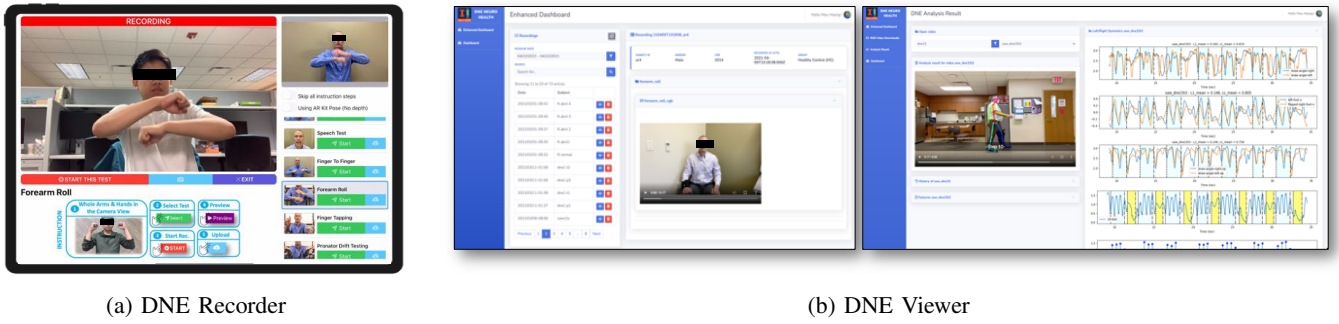


Fig. 2: DNE System. (a) *DNE Recorder* - an iOS application for neurological recordings collection. (b) *DNE Viewer* a web application for dataset management, video previewing and visualizing the analysis results.

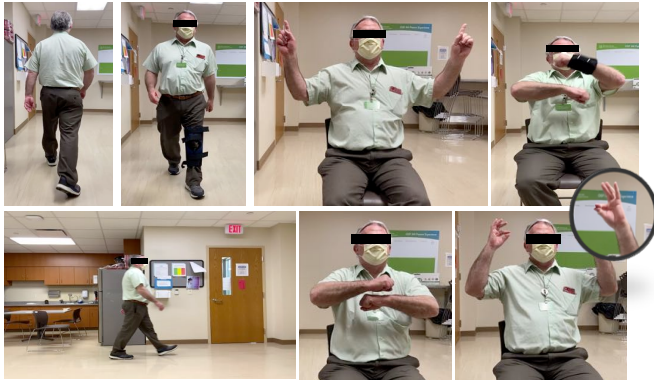


Fig. 3: Examples of DNE dataset recordings. Impairments are induced by wearing a wrist brace for FR, a rubber band for FT and a knee brace for SAW tests.

KIMORE provides RGB, depth, and pose data for each recording, collected by Kinect v2 which is not as ubiquitous as handheld devices adopted in DNE. In Table II, we compare our dataset versus state of the art public gait impairment datasets in various aspects. For this comparison, we only focus on studies using a single-view, portable camera for data collection, similar to our setting. Accordingly, we list the contributions introduced by our dataset as: 1) This is the first public dataset studying multiple neurological test segments. 2) Our dataset includes normal and abnormal performance of the same task for each particular subject. This opens an opportunity for detailed personalized neurological health assessment and longitudinal data analysis, 3) our dataset contains multiple data modalities, including depth videos, camera parameters and 2D/3D pose estimation. Details will be described further in Section V-A.

V. DNE VISION-BASED ANALYSIS

In our DNE analysis pipeline, given an RGB video, we first compute the human pose in each frame. Next, from the pose time series, we extract a set of features that quantify the subject's performance in various aspects. We structure our analysis pipeline into three layers, namely 1) pose estimation, 2) feature extraction and 3) application layer, as illustrated in Fig. 4. The pose estimation layer provides frame-level high-quality 2D/3D joint locations using state-of-the-art

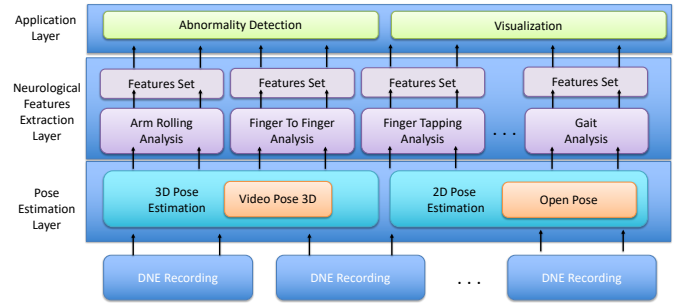


Fig. 4: Overview of DNE video analysis framework.

frameworks (Section V-A). We pre-process the estimated pose to prepare it for feature computation. In the feature extraction layer, we calculate a set of features that describe subject's performance on the test. We carefully design these features for each test segment separately to accurately reflect the subjects performance and any abnormality. Lastly, the application layer contains several downstream tasks consuming the features set, including abnormality detection and visualization for a qualitative comparison between recordings. In this section, we elaborate upon the aforementioned steps in detail.

A. Pose estimation

For upper body tests including FT, FR and FTF, we use OpenPose (OP) [18] to estimate the 2D pose. On the other hand, for SAW tests, we compute the 3D pose using the VideoPose3D package [19].

OpenPose (OP) [18], [59] is designed for multi-person real time 2D hand and body pose estimation. For body pose estimation and given an RGB image, OP first detects all visible body parts in the image and the corresponding part affinity fields. Associating body parts to each individual in an image boils down to a graph matching problem which is broken into smaller subproblems and solved in a greedy fashion. OP package also provides a 2D hand keypoint detector based on [59]. This hand keypoint detector is trained using a process called multi-view bootstrapping to further enable robustness to hand occlusions inherent in a single-view image.

VideoPose3D (VP3D) [19] estimates 3D pose of a single person given a video. VideoPose3D adopts dilated temporal convolution over the sequence of 2D keypoints extracted

TABLE II: A comparison between multiple vision-based gait impairment video datasets, acquired by a single camera.

Dataset	Availability	Sagittal View	Frontal View	Data Type	Mobile Device	Number of Subjects	Number of Sequences	Pose Estimation	Normal and Abnormal Pairs
Xue et al. [15]	✗	-	-	RGB	✗	-	-	2D	✗
Sato et al. [55]	✗	✗	✓	RGB	✗	2	2	2D	✗
Ortells et al. [42]	✓	✗	✓	Binary	✗	10	20	✗	✓
Nieto-Hidalgo et al. [43]	✓	✓	✓	Binary	✓	-	73	✗	✓
Kidzinski et al. [57]	✓	✓	✗	RGB	✗	1026	1792	2D	✗
Ours	✓	✓	✓	RGB/D	✓	21	336	2D/3D	✓

from the video frames. Thus, it operates on a subset of 2D poses extracted from overlapping segments of the video. The main assumption here is the smoothness of human's motion. Therefore, at any timepoint, processing a short sequence of 2D keypoints before and after the designated timepoint leads to an accurate and less noisy pose estimation. The 3D pose in VideoPose3D is normalized such that the location of each keypoint is relative to a root joint such as pelvis. VideoPose3D is trained on public datasets including Human3.6m [60] and HumanEva-I [61].

For upper body tests, if the subject and the moving limb is located parallel to the camera plane, then the motion is well approximated in a plane, i.e. in two dimensions. That is why we chose to extract 2D pose data for upper body tests. However, this might not hold for the SAW test (especially depending in the camera view), hence urging us to use 3D pose for this analysis.

B. Pre-processing

We truncate a recording to only include the sequence of frames that are related to the subject performing the test. Specifically, for SAW, the subjects are asked to walk the length of a room multiple times. We partition out the turning segments and exclude them in our analysis.

To account for variable distance of the subjects from the camera, we normalize the estimated pose by a reference length. For FT, FR and FTF tests, the reference is the length of the forearm. For SAW, the reference length is the distance between the pelvis and neck joints. We compute the reference lengths as the median of the value across all the frames. We choose median as it is a more robust statistics to outliers. In addition, as the estimated pose can be erroneous at some frames we use median and Savitzky-Golay filtering [62]. In our dataset, we have excluded 20 number of recordings for which the pose estimation is unreliable and exhibits significant noise. Therefore, we only analyzed 334 videos in total.

C. Notations

Given the pose sequence estimated from the RGB video, we extract a set of quantified features. Below, we first express our notations and then introduce the features we defined for each test.

Let $v = [v_1, \dots, v_N]$ denote the set of N frames ordered chronologically in video v . There is a one-to-one correspondence between the time associated with each frame and the frame index, where $t = [t_1, \dots, t_N]$ and $t_i = i/fps$, fps denoting the frame per second rate of the video. Given v and

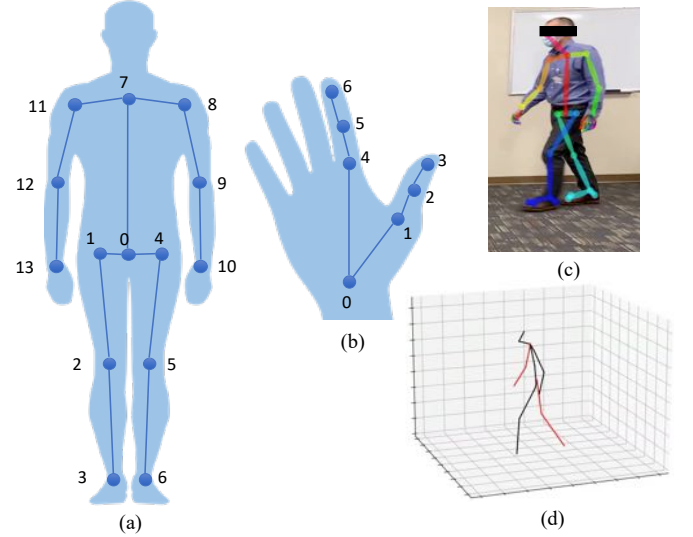


Fig. 5: Skeleton tree for (a) body B and (b) hand H. Examples of human pose estimation (c) in 2D (B_2, H_2) using OpenPose [18] and (d) in 3D (B_3) using VideoPose3D [19].

the pose estimation module (such as OP or VP3D), we extract the location of K keypoints in each frame. For convenience, we use the same indexing of the body joints for both 2D and 3D pose. However, to differentiate between the 2D and 3D pose, we denote each by B_2 and B_3 , respectively. Furthermore, we use H_2 to represent the 2D hand keypoints. An illustration of the hand and body skeleton trees alongside our indexing notations are provided in Fig. 5. Note that, for the sake of brevity, we have only indexed a subset of the keypoints that we are using in our analysis.

We reserve $s_{k,*}[i]$ for the location of the k -th keypoint at frame i , corresponding to skeleton tree $* \in \{H_2, B_2, B_3\}$. For $* \in \{H_2, B_2\}$, $s_{k,*}[i] \in \mathbb{R}^2$ and for $* = B_3$, $s_{k,*}[i] \in \mathbb{R}^3$. Furthermore, we add superscript r and l to point to right and left (R/L) body parts, respectively. For example, $s_{3,H_2}^r[i]$ marks the location of the tip of the right thumb at frame i .

To extract kinematic features that quantify the performance of a subject in a test, we track the location of various major keypoints and define a set of features accordingly. Major keypoints vary based on the test. For instance, the major keypoints in FT include the tip of the index and thumb fingers of right and left hands while in FR, we closely track the wrist joints.

In different tests, the subjects are asked to move certain limbs repeatedly. This implies that the movement of the subject

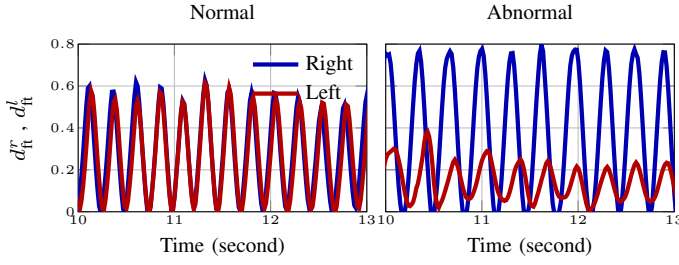


Fig. 6: FT amplitude for normal and abnormal examples.

and the location of a subset of the keypoints follows a periodic pattern. Thus, it is natural to compute features such as period, frequency and amplitude for periodic pose patterns and report the mean and standard deviations across different cycles. In addition, for a test performed normally, the features (such as period, frequency and amplitudes) corresponding to the right and left body parts should be close. Thus, to quantify the difference between the right f^r and left f^l features, we define an asymmetry metric as:

$$\text{Asym}(f^r, f^l) = \frac{|f^r - f^l|}{f^r + f^l}. \quad (1)$$

Another useful metric in our analysis is Pearson correlation coefficient denoted by CC. For two 1D discrete time series x_1 and x_2 , we define CC as:

$$\text{CC}(x_1, x_2) = \frac{(x_1 - \bar{x}_1)^T (x_2 - \bar{x}_2)}{\|x_1 - \bar{x}_1\|_2 \|x_2 - \bar{x}_2\|_2}. \quad (2)$$

where $\bar{\cdot}$ and \cdot^T denote the mean and transpose operators. Contextually, if x_1 and x_2 are the positions of two joints, $\text{CC}(x_1, x_2)$ reflects the similarity in their trajectories. For highly correlated series, $|\text{CC}|$ is close to one.

D. Feature definition

We list the features defined for various tests in Table III and describe them in detail below.

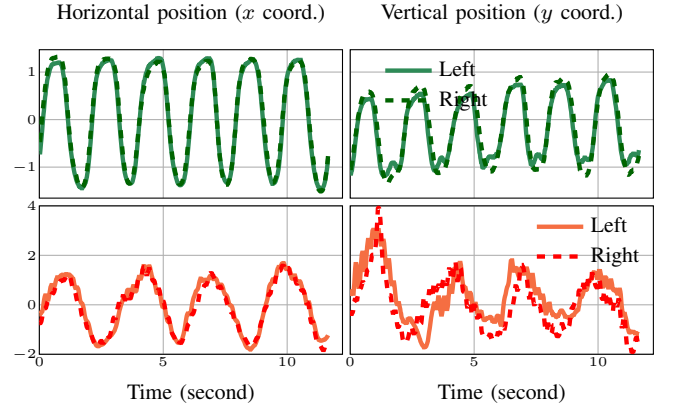
Finger Tapping (FT): For this test, the major keypoints are the tip of the R/L thumb and index fingers alongside R/L wrist and elbow joints. To extract properties of the periodic motion, we look at the distance between the tip of the index and thumb fingers across time defined as:

$$d_{ft}^*[i] = \|s_{3,H_2}^*[i] - s_{6,H_2}^*[i]\|_2, \quad * \in \{r, l\}. \quad (3)$$

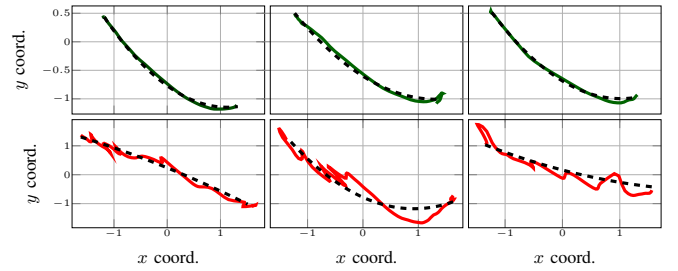
Examples of d_{ft}^r and d_{ft}^l for normal and abnormal executions of the FT test are provided in Fig. 6. Note that in our dataset, to simulate abnormality in FT the subjects are wearing a rubber band around index and thumb fingers of one hand. As also revealed by Fig. 6, this limits the tapping amplitude of the hand wearing the band and slows down the tapping rate.

Given d_{ft}^* , we compute the period for the $*$ hand, T_{ft}^* , as the time (in seconds) between two consecutive local minima (or maxima) of d_{ft}^* . Frequency F_{ft}^* is the reciprocal of T_{ft}^* . We also report the magnitude of finger-tapping A_{ft}^* as the difference in consecutive minimum and maximum values of d_{ft}^* .

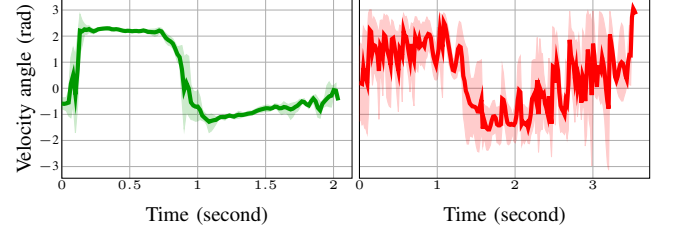
Furthermore, to evaluate the stability of the hands and arms during the FT recording, we examine the wrist and elbow joints. For this purpose, we introduce the relative height



(a) Horizontal (first column) and vertical (second column) locations of joint index 5 in H_2 for normal (first row) and abnormal (second row). Within each row (column), the subplots share the same vertical (horizontal) axis.



(b) Actual trajectory (solid line) vs. smoothed fitted trajectory (dashed line). The subplots within each row share the same vertical axis.



(c) Cycle-wise average of velocity angle. The subplots share the same vertical axis.

Fig. 7: Examples of FTF features including (a) finger positions, (b) finger movement trajectory, (c) angle of the velocity. Green (red) curves stand for normal (abnormal) recordings.

between (s_{10,B_2}, s_{13,B_2}) and (s_{9,B_2}, s_{12,B_2}) across N frames, in other words:

$$S_{ft}^{\text{wrist}} = \frac{1}{N} \sum_{i=1}^N \frac{\|s_{10,B_2}[i] - s_{13,B_2}[i]\|_2}{\|s_{13,B_2}[i]\|_2}, \quad (4)$$

$$S_{ft}^{\text{elbow}} = \frac{1}{N} \sum_{i=1}^N \frac{\|s_{9,B_2}[i] - s_{12,B_2}[i]\|_2}{\|s_{12,B_2}[i]\|_2}. \quad (5)$$

We also report the asymmetry of the period ($\text{Asym}(T_{ft}^r, T_{ft}^l)$), frequency ($\text{Asym}(F_{ft}^r, F_{ft}^l)$) and amplitudes ($\text{Asym}(A_{ft}^r, A_{ft}^l)$) of R/L hands following (1). Furthermore, we define the instant tapping speed and acceleration for R/L hands as the first and second order derivatives of d_{ft}^r and d_{ft}^l with respect to time. We adopt mean and maximum of instant speed and acceleration

TABLE III: Summary of our DNE features. Asymmetry between R/L features is computed based on (1).

Finger Tapping (FT)	Finger to Finger (FTF)	Forearm Roll (FR)	Stand and Walk (SAW)
Amplitude <i>Mean, STD, Median, Asymmetry</i> Maximum distance between the tip of the index and thumb fingers	Horizontal symmetry <i>CC</i> The spatial trajectory of the R/L index finger joint in the horizontal direction	Amplitude <i>Mean, STD, Median, Asymmetry</i> Distance between the minima and maxima of the horizontal position of the R/L arms	Knee angle symmetry <i>Mean, STD, Median</i> The CC of the aligned knee angle series on the R/L legs within a walking segment (a full pass with the room length)
Period <i>Mean, STD, Median, Asymmetry</i> Time taken to complete one tapping cycle for R/L hands	Vertical symmetry <i>CC</i> The spatial trajectory of the R/L index finger joint in the vertical direction	Period <i>Mean, STD, Median, Asymmetry</i> Time taken to complete one forearm roll cycle for R/L hands	Step symmetry <i>Mean, STD, Median</i> The CC of the aligned spatial trajectory of the R/L foot in the horizontal direction
Maximum speed (R/L) <i>Mean, Asymmetry</i> Maximum of instant tapping speed (defined as the derivative of the distance between the tip of the index and thumb fingers with respect to time) for R/L hands	Period (R/L) <i>Mean, STD</i> Total time taken for one complete cycle (moving from the highest to the lowest position and back) on each side	Maximum speed (R/L) <i>Mean, STD, Median, Asymmetry</i> Maximum of forearm roll speed (defined as the first derivative of the vertical coordinate of the wrist joint with respect to time) for R/L hands	Step length <i>Mean, STD, Median</i> The furthest distance between two feet at each step cycle
Maximum acceleration (R/L) <i>Mean, STD, Median, Asymmetry</i> Maximum of instant tapping acceleration (defined as the second derivative of the distance between the tip of the index and thumb fingers with respect to time) for R/L hands	Average speed <i>Mean, STD, Median</i> The traversed distance of R/L finger within half a cycle's period divided by half the cycle's period	Maximum acceleration (R/L) <i>Mean, STD, Median, Asymmetry</i> Maximum of forearm roll acceleration (defined as the second derivative of the vertical coordinate of the wrist joint with respect to time) for R/L hands	Step width <i>Mean, STD, Median</i> The shortest distance between two feet at each step cycle
Tapping speed (R/L) <i>Mean, STD, Median</i> Average tapping speed (defined as the maximum distance between the tip of the index and thumb fingers divided by half the tapping period), for R/L hands	Path smoothness (R/L) <i>Mean, STD, Median</i> The ratio between the traversed distance of R/L finger and the length of the fitted smooth curve to the actual trajectory	Rolling speed <i>Mean, STD, Median</i> Average forearm roll speed (defined as the amplitude divided by half the rolling cycle period) feature	Step time <i>Mean, STD, Median</i> The time to complete one step cycle (the interval between two consecutive time-points having the shortest distance between two feet)
Tapping rate (R/L) <i>Mean, STD, Median</i> The number of finger taps per second for R/L hands	Velocity angle symmetry (R/L) <i>Mean, STD</i> The pairwise CC between every combination of the angle velocity series of two cycles	Rolling rate Number of forearm roll cycles per second	Time to stand Total time taken from the first stand up effort to a full standing on feet state
Wrist stability <i>Mean, STD, Median</i> Variations in R/L wrist joint positions			Turning time <i>Mean, STD, Median</i> Total time taken for a subject to turn around after each walking segment
			Walking speed <i>Mean, STD</i> Total of traveled distance of the pelvis joint divided by the duration of a walking segment
			Cadence <i>Mean, STD</i> Total number of steps divided by the duration of a walking segment

across tapping cycles as a feature in our analysis. Additionally, we define the average tapping speed for R/L hands as the maximum of d_{ft}^r and d_{ft}^l divided by half the tapping period. Finally, we introduce tapping rate as the number of finger tappings per second.

Finger to Finger (FTF): Empirically, for our dataset, we observe that the estimated pose by OP for middle joints of the index finger, i.e. joint index 5 in H_2 , is more stable compared to the outer fingertip. Hence, we focus on this joint for FTF test. In a normal performance of FTF, the horizontal and vertical trajectories of the right and left hands are symmetric up to a mirroring (Fig. 7a- top row), while this does not necessarily hold for abnormal case (Fig. 7a- bottom row). Thus, in each cycle, we define the cross correlation between the horizontal (x) and vertical (y) coordinates as the horizontal ($S_{ftf}^{finger-x}$) and vertical symmetries ($S_{ftf}^{finger-y}$):

$$S_{ftf}^{finger-x} = CC([s_{5,H_2}^l]_x, -[s_{5,H_2}^r]_x), \quad (6)$$

$$S_{ftf}^{finger-y} = CC([s_{5,H_2}^l]_y, [s_{5,H_2}^r]_y). \quad (7)$$

where $[s_{5,H_2}^\dagger]_* = \{s_{5,H_2}[i]\}_{i=1}^N$, $* \in \{x, y\}$ and $\dagger \in \{r, l\}$, is the x or y coordinates of the pose series. We also compute the period and average speed. We derive the average speed by dividing the traversed distance of R/L finger within half a

cycle's period by half the cycle's period.

Patients with neurological impairments tend to have tremors while moving their fingers during FTF test [63]. This leads to a deviation of the fingers' trajectory from a smooth curve. To characterize this deviation, we first fit a smooth curve to the fingers' trajectory, in the form of a second order polynomial in terms of the vertical and horizontal coordinates. Heuristically, we observe that fitting a second order function to the trajectories, well matches the FTF trajectories of normal subjects. We consider the length of this smooth curve as a reference to compare against the length of the original fingers' trajectory. We then define the ratio of the length of the actual fingers' trajectory during each FTF cycle by the length of the fitted smooth curve as path smoothness metric (PS). We report PS for both right and left hands. Examples of normal and abnormal finger trajectories alongside the smooth fitted curves are plotted in Fig. 7b.

Another feature we found helpful in detecting abnormal function in FTF is instant velocity. We derive instant velocity by taking the first derivative of the horizontal and vertical pose with respect to time. We then examine the slope of the instant velocity vector for right and left hands separately, by computing the angle between its vertical and horizontal components. Therefore, at each time instant t , we define

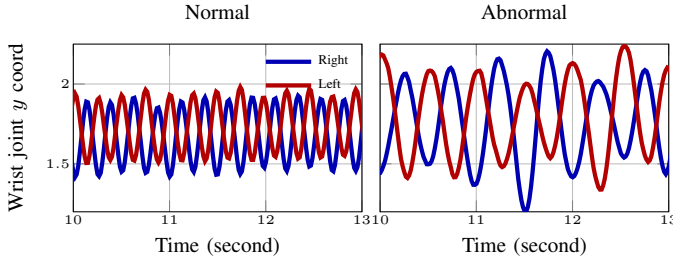


Fig. 8: Vertical (y) coordinate of the wrist joint versus time for two normal and abnormal examples in FR test.

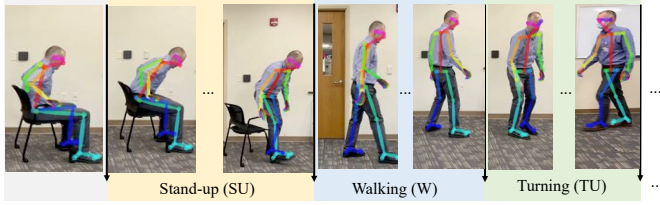


Fig. 9: An example of non-overlapping segments of the SAW test.

velocity angle as:

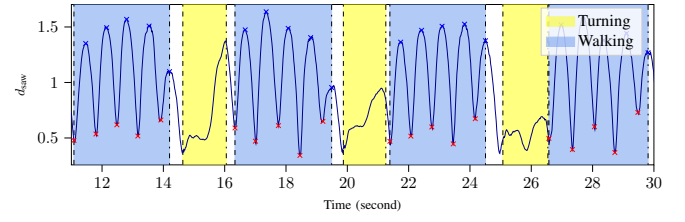
$$\theta^*(t) = \text{atan2} \left(\frac{d[s_{5,H_2}^*]_y}{dt} \right), * \in \{r, l\}. \quad (8)$$

Next, for each hand, we compare θ across different cycles using CC metric in (2). Given N_C number of cycles, we have $\binom{N_C}{2}$ CC values assessing the symmetry of the R/L velocity angles across different cycles, which we summarize by reporting the mean and standard deviation. Examples of normal and abnormal aligned velocity angles across different cycles are provided in Fig. 7c. Note that for abnormal FTF, large magnitude fluctuations visibly appear in the velocity angle series across different cycles, caused by tremors in moving the hands.

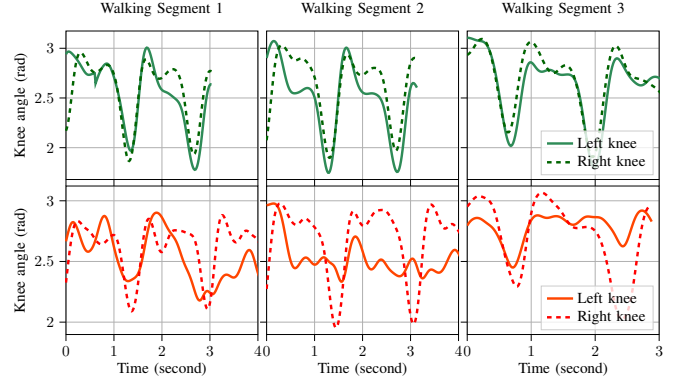
Forearm Rolling (FR): We include the wrist and elbow joints as the major keypoints for this test. We specifically attend to the vertical coordinate of the wrist joints to compute period T_{fr}^* , frequency F_{fr}^* and amplitudes A_{fr}^* for $* \in \{r, l\}$. Fig. 8 illustrates the vertical position of the R/L wrists for a normal and abnormal example. Note that, due to wearing the device in the abnormal recording, the period of the forearm roll cycles for both R/L hands are larger compared to its normal counterpart. In addition, similar to FT, we include the asymmetry of the aforementioned metrics in the FR feature set.

In addition, we include the maximum instant speed and acceleration derived from vertical coordinates of the wrist joints. Similar to FT, we define rolling speed and rate. Rolling speed is computed as the difference between the minimum and maximum of y coordinate of the R/L hands divided by half the rolling period. Also, rolling rate is defined as the number of rolling cycles per second. Finally, we report the stability of the elbows S_{fr}^{elbow} and define it analogous to (5).

Stand and Walk (SAW): We use the side-view SAW recordings in our analysis of SAW test. As stated earlier, for



(a) Distance between two feet (d_{saw}) versus time. Marker \times denotes the starting and \times denotes the ending of each step. In this example, a SAW video is partitioned into 4 walking (W) and 3 turning (TU) segments.



(b) Knee angle series for three walking segments taken from a normal and an abnormal recording of the same subject. Green (red) curves stand for normal (abnormal) recordings.

Fig. 10: Representative examples of SAW features.

SAW pose estimation we use VP3D [19]. In VP3D, the joint locations are defined relative to the pelvis joint. As a result, estimated pose by VP3D misses the global position of subjects within a frame which is essential to detect different segments of the SAW test, i.e. stand-up (SU), walk (W) and turn (TU). This urged us to track the 2D position of the pelvis joint s_{0,B_2} extracted by OP as a notion of subject's global position in a video frame. Analyzing this position through time enables us to split a SAW recording into multiple non-overlapping SU, W and TU segments. Fig. 9 illustrates an example of these segments.

For the SU segment of SAW, we focus on the time to stand [64], measured by the total time taken from the first SU effort to a full standing on feet state. We derive time to stand by thresholding the magnitude of the pelvis joint's velocity. Note that, since our subjects are asked to walk back and forth a designated room multiple times, at some points, they have to change direction and turn around. We report time to turn around as another indicative feature for SAW test.

The first set of features derived for the walking phase of SAW are obtained based on the distance between the two feet stated as:

$$d_{\text{saw}}[k] = \|s_{3,B_3}[k] - s_{6,B_3}[k]\|_2.$$

Note that the periodic nature of a normal gait also reflects in d_{saw} (see Fig. 10a). Furthermore, given d_{saw} , we highlight different W and TU segments in Fig. 10a. For a gait pattern derived based on the distance between the two feet, i.e. d_{saw} ,

step time is the time to complete one step and computed as the time difference between two consecutive local maxima of d_{saw} . Meanwhile, *step length* defined as linear distance between two successive placements of the same foot [65] manifests as the local maxima of d_{saw} . The *step width*, on the other hands, is interpreted as the local minima of d_{saw} . In our calculations of the step length, step width and step time, we exclude the turning segments and report mean and STD statistics of these metrics.

As two global features for gait, we report mean and STD of *cadence* and *average speed* across all W segments. We compute cadence as the number of steps divided by the duration of a walking segment. In addition, average speed is determined by the total traveled distance of the pelvis joint divided by the duration of a walking segment.

To evaluate the symmetry between the right and left gait, we introduce the cross correlation between the knee angle series of R/L legs, denoted by $S_{\text{saw}}^{\text{knee angle}}$. We find this feature a good descriptive of gait abnormality as in our recordings, gait abnormality is introduced through wearing a knee band which limits the knee motion (Fig. 3). In every frame, we define the knee angle as the angle between $s_{5,B_3} - s_{4,B_3}$ and $s_{5,B_3} - s_{6,B_3}$ for the right leg and $s_{2,B_3} - s_{1,B_3}$ and $s_{2,B_3} - s_{3,B_3}$ for the left leg. As there is a lag between the R/L gait cycles, we align the knee angle series of the R/L legs within each cycle and then report CC of the aligned series. Examples of aligned normal and abnormal knee angles for R/L legs are shown in Fig. 10b. For normal gait, the R/L knee angles are highly correlated after alignment (Fig. 10b-top row), while this does not hold for abnormal gait (Fig. 10b-bottom row).

In addition, we define step symmetry between the right and left feet movements by comparing the horizontal position of R/L feet at different gait cycles. We represent this metric by $S_{\text{saw}}^{\text{feet-x}}$. To compute $S_{\text{saw}}^{\text{feet-x}}$, similar to $S_{\text{saw}}^{\text{knee angle}}$, we first align the R/L horizontal positions within each gait stride and report the CC of the aligned series. We report mean and STD for both $S_{\text{saw}}^{\text{feet-x}}$ and $S_{\text{saw}}^{\text{knee angle}}$ across different cycles.

VI. RESULTS AND DISCUSSION

A. Subject-based Normal vs. Abnormal Comparison

In this section, we aim to compare the normal and simulated-impaired performances of the same subject and show that this analysis is insensitive to the choice of recording device and robust to the viewpoint or distance from the camera. Note that in our dataset, for each subject, we have four sets of recordings. Two of these recordings capture the normal performance of the test, while in the other two, the subject is asked to perform abnormally. In addition, two pairs of normal/abnormal recordings are captured by an iPhone (P) and an iPad (T). Let N_T/N_P and A_T/A_P denote the normal and abnormal recordings captured by iPhone/iPad.

For each feature set and subject, we define A-A/N-N as the intra-class distance between the features derived from the abnormal/normal recordings of the subject captured on iPhone and iPad devices. In other words, A-A is the distance between features computed for A_T and A_P recordings, while N-N marks the difference between the features of N_T and N_P

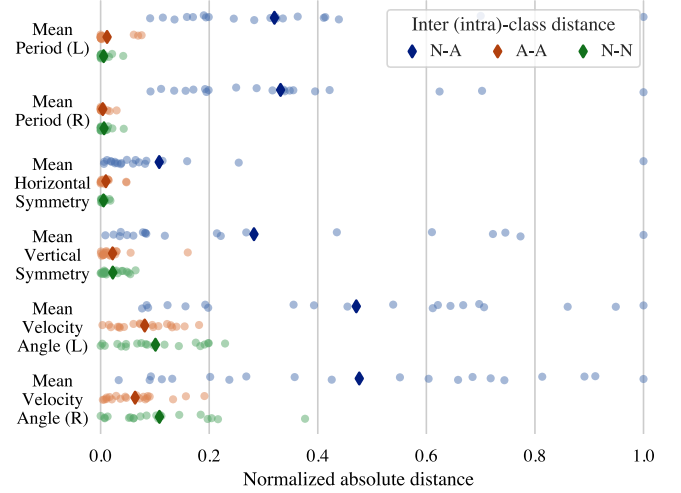


Fig. 11: The inter-class and intra-class distances between normal (N) and abnormal (A) recordings of some features, including mean period, horizontal asymmetry, vertical asymmetry, velocity angle asymmetry from the FTF test. All measurements are normalized to the unit interval. Marker \blacklozenge denotes the mean value.

videos. For N-A, we consider the distance between A_T-N_P and N_T-A_P pairs and report the average as the N-A inter-class distance for each subject. We normalize the A-A, N-N, and N-A distances by the maximum of N-A distances.

Fig. 11 illustrates the distribution of A-A, N-N, and N-A distances across 20 different subjects for a subset of features. While the intra-class values are concentrated near zero, the inter-class distances are spread out over a wider range. In addition, the mean A-A and N-N distances are strictly lower than the N-A distances. The higher concentration of A-A and N-N distances around zero shows that our feature set is robust to some minor changes in the viewpoint and is not affected by the recording device. Furthermore, it can be seen as a proof-of-concept, demonstrating the ability to compare the subject's performance across different time points. This validates the potential of tracking the progression of neurological disease and the effects of treatment over time using our DNE system and features.

B. Abnormality Detection

PCA Analysis: The feature set describing normal and abnormal recordings constitutes a high-dimensional vector. For a visual comparison of normal and abnormal recordings in terms of their derived features, we perform dimensionality reduction through principal component analysis (PCA). For this purpose, for each test, we concatenate the set of features listed in Table III and normalize them before passing them to PCA. Fig. 12 showcases the PCA results for different tests. It is observed that the normal and abnormal recordings are separated in dimension reduced feature space. This implies that the defined features are descriptive of normal and abnormal performance.

Abnormality Detection: We assess the normal and abnormal

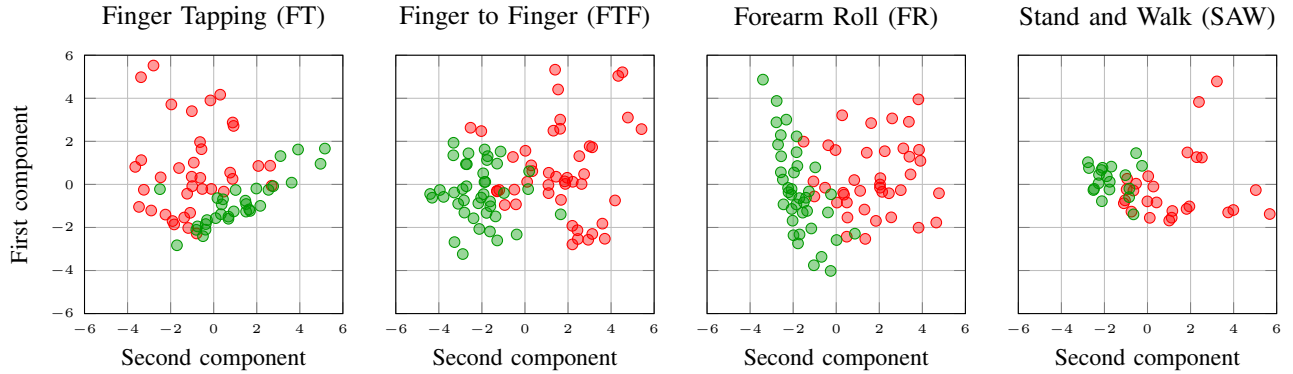


Fig. 12: PCA analysis for different tests including FT, FTF, FR and SAW. Green (red) circles stand for normal (abnormal) recordings. All subplots share the same vertical axis.

classification performance using our proposed set of features. For this purpose, we utilize several machine learning (ML) models that are grouped into: 1) tree-based methods such as Decision Tree (DT), Random Forest (RF), Gradient-Boosting Machine (GBM) [66], XGBoost [67] and 2) parametric models trained using gradient-descent updates, including Logistic Regression (LR), Support Vector Machine (SVM) and Multi-layer Perceptron (MLP). The architecture of our MLPs trained on FT, FTF, and FR datasets include three fully connected layers with 100-dimensional hidden layers. For the SAW test, the MLP architecture contains four fully connected layers with hidden layers of length 80, 200, and 80. The non-linearity used in our MLPs is rectified linear unit (ReLU). To evaluate different models, we use metrics such as accuracy, average precision, F1 score, and area under the curve (AUC).

We have two splitting schemes to separate the train from test sets in our experiments. In *video-based* splitting, videos from all subjects are divided independently based on a 80%/20% splitting ratio for train/test sets. In addition, to evaluate the performance of the models on unseen patients, the *subject-based* division scheme splits a portion of the subjects into the train set while keeping the rest in the test set. Thus, videos belonging to the subjects in the train set are not used in the test set and vice versa. In subject-based splitting, we have 16/4 subjects in train/test sets.

Table IV summarizes the average classification performance of the ML models trained on our DNE feature set. While all models perform well for various tests, among gradient-based approaches, MLP and LR models produce the most competitive results, while RF is the best candidate among tree-based methods. For FT, FR, and SAW tests, gradient-based methods outperform tree-based ones. However, the gap between the two groups of models is not significant across different tests. This suggests that the extracted set of features well-distinguish normal from abnormal cases. This enables a wide range of ML models to perform well on the binary classification task using these features.

C. Feature Importance Analysis

One benefit of tree-based models is in the tractable decision-making process. Therefore, we investigate the importance of

each feature, contributing to the decision process by analyzing our random forest models. This analysis gives us the weight of all features, which we sort in descending order in Fig. 13.

We notice that symmetry between specific R/L features for FT, FTF, and SAW tests is considered the most important, i.e., with the largest weight. For the SAW test, the most important feature is the similarity between the knee angle time series across different cycles ($S_{\text{SAW}}^{\text{knee angle}}$) while for FT (13a) and FTF (13b), the features with the largest weights are instant speed and fingertip position ($S_{\text{FTFinger-x}}$) symmetries, respectively. Although this can be attributed to the nature of the simulated impairments in our dataset, it is consistent with the clinical practice, where the left and right asymmetry is a common biomarker [68]–[70] of different neurological disorders.

Furthermore, temporal and spatial features that characterize the periodic behavior of the movement are important metrics that the decision tree classification models rely on to discriminate normal from abnormal. Examples of these features are amplitude and period for FT, FTF, and FR tests, step length, width, and step time for SAW.

We also notice that for a subset of features, having large variations (i.e. STD) across different cycles is another indicator of abnormal performance in our dataset. This is captured in the large weight associated with STD values of some features for various tests.

D. Discussion

In this work, we emphasized the design of interpretable features that enables a reliable classification between normal and abnormal performances. Having powerful digital bio-markers further reduces the workload of classification models and improves their generalization, especially when large datasets are not available for training. In addition, unlike black-box deep-learning-based models, the explainability introduced by our diverse set of features allows clinicians to better understand and track patients' status through time.

Note that natural motion properties differ across various subjects. For example, one subject can be inherently slower or have less strength in performing some tests. In our dataset, we witnessed while some subjects had a slower inherent speed in their normal performance, they were mistakenly classified

TABLE IV: Classification performance of several machine learning models for various tests including FT, FTF, FR and SAW. The best and second best results are highlighted in **bold** and underline, respectively.

Test	Model	Subject Based						Video Based					
		Acc	Precision	Recall	AP	AUC	F1 Score	Acc	Precision	Recall	AP	AUC	F1 Score
SAW	DT	0.6986	0.7638	0.7317	0.6530	0.7033	0.6832	0.6889	0.7133	0.6967	0.6083	0.6983	0.6543
	GBM	0.7288	0.7595	0.7933	0.6740	0.7383	0.7322	0.7111	0.7190	0.8033	0.6363	0.7350	0.7100
	LR	<u>0.8163</u>	0.8367	<u>0.8100</u>	<u>0.7744</u>	<u>0.8225</u>	<u>0.8012</u>	0.7778	0.7533	0.7700	0.7003	0.7867	0.7476
	MLP	0.8792	0.8850	0.8800	0.8336	0.8800	0.8677	0.8667	0.8600	0.8600	0.8072	0.8683	0.8529
	R.Forest	0.8000	0.8333	0.7850	0.7367	0.7958	0.7789	<u>0.8000</u>	<u>0.8133</u>	0.7367	<u>0.7033</u>	<u>0.7917</u>	<u>0.7544</u>
	SVM	0.7927	<u>0.8550</u>	0.7517	0.7543	0.8017	0.7653	0.7556	0.8000	0.7200	<u>0.7033</u>	0.7700	0.7167
	XGBoost	0.7232	0.7667	0.7600	0.6644	0.7300	0.7089	0.7333	0.7333	<u>0.8033</u>	0.6506	0.7517	0.7233
FTF	DT	0.9101	0.8889	0.9375	0.8653	0.9132	0.9072	0.9327	0.8944	0.9500	0.8722	0.9389	0.9176
	GBM	0.9056	0.8819	0.9250	0.8590	0.9083	0.8995	0.9111	0.8583	0.9500	0.8444	0.9167	0.8976
	LR	<u>0.9156</u>	<u>0.9096</u>	<u>0.9292</u>	<u>0.8756</u>	<u>0.9215</u>	<u>0.9103</u>	<u>0.9438</u>	<u>0.8992</u>	0.9833	<u>0.8937</u>	<u>0.9556</u>	<u>0.9356</u>
	MLP	0.9094	0.8838	0.9542	0.8638	0.9153	0.9099	0.9438	0.8992	0.9833	0.8937	<u>0.9556</u>	<u>0.9356</u>
	R.Forest	0.9389	0.9399	<u>0.9417</u>	0.9164	0.9417	0.9376	0.9778	0.9714	0.9833	0.9659	0.9833	0.9759
	SVM	0.8625	0.8578	0.8708	0.8058	0.8653	0.8542	0.8876	0.8429	0.8917	0.8048	0.8931	0.8630
	XGBoost	0.9056	0.8849	0.9083	0.8582	0.9083	0.8936	0.9111	0.8643	0.9167	0.8429	0.9167	0.8858
FR	DT	0.9508	0.9206	0.9833	0.9101	0.9513	0.9455	0.9350	0.8911	0.9667	0.8703	0.9360	0.9195
	GBM	0.9638	0.9417	0.9833	0.9313	0.9659	0.9566	0.9608	0.9333	0.9667	0.9125	0.9652	0.9418
	LR	0.9867	0.9700	1.0000	0.9700	0.9871	0.9836	0.9733	0.9400	1.0000	0.9400	0.9742	0.9673
	MLP	<u>0.9742</u>	0.9533	1.0000	<u>0.9533</u>	<u>0.9705</u>	<u>0.9745</u>	0.9733	0.9400	1.0000	0.9400	0.9742	0.9673
	R.Forest	0.9633	0.9317	1.0000	0.9317	0.9659	0.9605	0.9600	0.9133	1.0000	0.9133	0.9652	0.9495
	SVM	0.9596	<u>0.9589</u>	0.9542	0.9339	0.9580	0.9529	0.9483	0.9400	0.9417	0.9067	0.9451	0.9357
	XGBoost	0.9638	0.9417	0.9833	0.9313	0.9659	0.9566	0.9608	0.9333	0.9667	0.9125	0.9652	0.9418
FT	DT	0.7535	0.7316	0.7530	0.6620	0.7517	0.7246	0.7670	0.7114	0.8143	0.6721	0.7694	0.7496
	GBM	0.7838	0.7494	0.8107	0.7045	0.7853	0.7677	0.8242	0.7781	0.8714	0.7388	0.8265	0.8111
	LR	<u>0.8934</u>	0.8566	<u>0.9357</u>	<u>0.8356</u>	<u>0.8974</u>	<u>0.8894</u>	0.8835	0.8584	<u>0.9214</u>	0.8339	0.8864	0.8842
	MLP	0.8987	<u>0.8617</u>	0.9482	0.8421	0.9036	0.8962	0.9275	<u>0.9100</u>	0.9464	0.8881	0.9299	0.9241
	R.Forest	0.8904	0.8875	0.8732	0.8324	0.8899	0.8723	0.9275	0.9250	0.8964	<u>0.8768</u>	<u>0.9215</u>	<u>0.9079</u>
	SVM	0.8816	<u>0.8617</u>	0.9006	0.8286	0.8870	0.8724	0.9132	<u>0.9100</u>	0.9179	0.8738	0.9156	0.9061
	XGBoost	0.7678	0.7605	0.7655	0.6930	0.7698	0.7407	0.7956	0.7781	0.8143	0.7102	0.7980	0.7720

as abnormal. This highlights the importance of taking into account the history of a subject and self-baselining. In our experiments, we showcased an example of self-comparisons of normal and abnormal performance of the same subject (Fig. 11). The purpose of this study was to show the ability of our designed features to discriminate between the varying status of the subject at different test times. This result validates the potential of our DNE pipeline as a personalized medical assessment system that helps clinicians better track and monitor the disease progression and the effect of medical treatments.

In our current DNE system, to have compatibility with iOS devices without LiDAR, we only operate on RGB videos without any calibrations. Thus, the absolute value of the spatial features such as amplitude in FT and FR or step length in SAW is not available. However, this is not the case for temporal or similarity/symmetry-based features. While it is beneficial to have the absolute value of the spatial features, in our current study we found that carefully normalized spatial features are still valuable in distinguishing normal from abnormal performance.

We view our system as one of the first attempts towards a multi-test pure vision-based DNE approach. Due to our modular design, our system can be easily extended to other tests, such as eye movement [71], facial activation [72], [73], or phonation [74]. Besides, enlarging the DNE dataset with recordings from real patients with different quantified severity levels can further validate the use of our current system in a

real-world setting. This would be the focus of our future work.

VII. CONCLUSION

In this paper, we proposed a comprehensive vision-based teleneurology solution named digital neurological examination (DNE). Using DNE software, users video record their performance on various motor tasks, including finger tapping, finger to finger, forearm roll, and stand-up and walk. We introduced the DNE dataset, a total of 336 videos consisting of normal and impaired functions of 21 subjects, performing different tests.

For each recording, 2D/3D pose is estimated and used to quantify kinematic and spatio-temporal features. These features form a set of digital biomarkers that can be 1) accurately obtained from common RGB videos with minimal calibration, 2) used to track the clinical changes across recordings at different time points.

On our DNE dataset, we analyzed the effectiveness of the defined features in differentiating normal versus impaired simulated videos per and across subjects. Our results demonstrate high accuracy and F1 scores for this classification using a variety of machine learning models. Future work might extend the setting of this study to a larger set of subjects with a diverse range of abnormalities.

REFERENCES

- [1] V. L. Feigin *et al.*, “Burden of neurological disorders across the us from 1990-2017: A global burden of disease study.” *JAMA neurology*, vol. 78, no. 2, pp. 165 – 176, 2021.

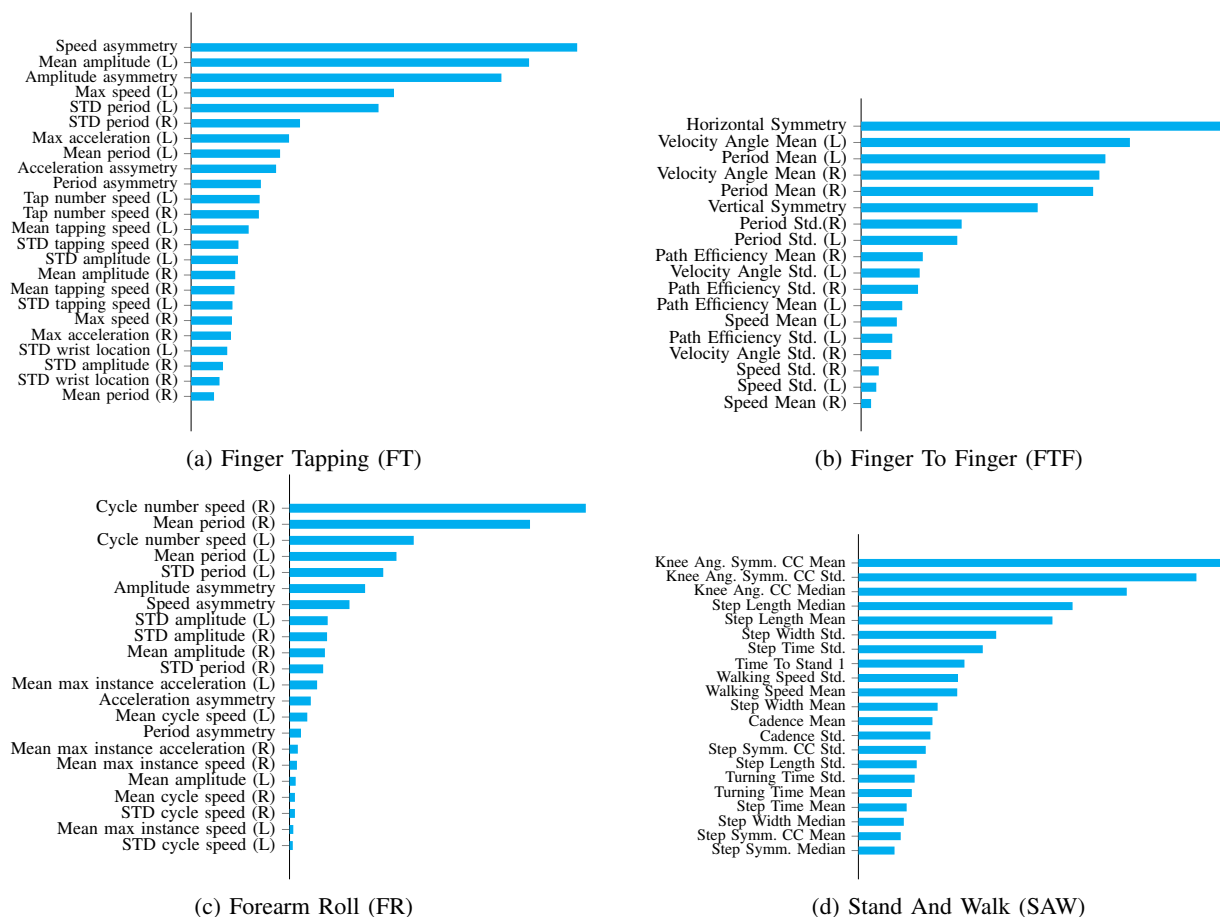


Fig. 13: Analyzing the importance of different features in our trained random forest models for various tests, including (a) FT, (b) FTF, (c) FR, and (d) SAW.

- [2] T. Dall *et al.*, "Supply and demand analysis of the current and future us neurology workforce," *Neurology*, vol. 81, no. 5, pp. 470–478, 2013.
- [3] S. N. Grossman *et al.*, "Rapid implementation of virtual neurology in response to the COVID-19 pandemic," *Neurology*, vol. 94, no. 24, pp. 1077–1087, Jun. 2020, publisher: Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology Section: Contemporary Issues.
- [4] M. Al Hussona *et al.*, "The Virtual Neurologic Exam: Instructional Videos and Guidance for the COVID-19 Era," *The Canadian Journal of Neurological Sciences. Le Journal Canadien Des Sciences Neurologiques*, pp. 1–6, May 2020.
- [5] D. Lupton *et al.*, "Telemedicine and the senses: a review," *Sociology of Health and Illness*, vol. 39, no. 8, pp. 1557–1571, 2017.
- [6] L. R. Wechsler, "Advantages and Limitations of Teleneurology," *JAMA Neurology*, vol. 72, no. 3, pp. 349–354, 03 2015.
- [7] R. Capozzo *et al.*, "Telemedicine is a useful tool to deliver care to patients with Amyotrophic Lateral Sclerosis during COVID-19 pandemic: results from Southern Italy," *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*, vol. 21, no. 7-8, pp. 542–548, Nov. 2020.
- [8] C. Duncan *et al.*, "Video consultations in ordinary and extraordinary times," *Practical Neurology*, vol. 20, no. 5, pp. 396–403, Oct. 2020, publisher: BMJ Publishing Group Ltd Section: How to do it.
- [9] V. Patterson, "Neurological telemedicine in the COVID-19 era," *Nature Reviews Neurology*, vol. 17, no. 2, pp. 73–74, Feb. 2021.
- [10] S. A. Mutgi *et al.*, "Emerging subspecialties in neurology: Telestroke and teleneurology," *Neurology*, vol. 84, no. 22, pp. e191 – e193, 2015.
- [11] E. R. Dorsey *et al.*, "Teleneurology and mobile technologies: the future of neurological care," *NATURE REVIEWS NEUROLOGY*, vol. 14, no. 5, pp. 285 – 297, 2018.
- [12] J. L. Adams *et al.*, "Digital technology in movement disorders: Updates, applications, and challenges," *Current neurology and neuroscience reports*, vol. 21, no. 4, pp. 16–16, 03 2021.
- [13] S. Williams *et al.*, "The discerning eye of computer vision: Can it measure parkinson's finger tap bradykinesia?" *Journal of the Neurological Sciences*, vol. 416, p. 117003, 2020.
- [14] R. Jaroensri *et al.*, "A video-based method for automatically rating ataxia," in *Proceedings of the 2nd Machine Learning for Healthcare Conference*, ser. Proceedings of Machine Learning Research, F. Doshi-Velez *et al.*, Eds., vol. 68. Boston, Massachusetts: PMLR, 18–19 Aug 2017, pp. 204–216.
- [15] D. Xue *et al.*, "Vision-based gait analysis for senior care," *CoRR*, vol. abs/1812.00169, 2018.
- [16] V. Tedim Cruz *et al.*, "A novel system for automatic classification of upper limb motor function after stroke: An exploratory study," *Medical Engineering & Physics*, vol. 36, no. 12, pp. 1704–1710, Dec. 2014.
- [17] M. Grobe-Einsler *et al.*, "Development of sarahome, a new video-based tool for the assessment of ataxia at home," *Movement Disorders*, vol. 36, no. 5, pp. 1242–1246, 2021.
- [18] Z. Cao *et al.*, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
- [19] D. Pavlo *et al.*, "3d human pose estimation in video with temporal convolutions and semi-supervised training," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [20] *Kinematics*. John Wiley & Sons, Ltd, 2009, ch. 3, pp. 45–81.
- [21] T. Ota *et al.*, "Forearm rolling test in precentral gyrus cerebral infarction," *Clinical Case Reports*, vol. 9, no. 6, p. e04201, 2021, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ccr3.4201>.
- [22] F. B. Horak *et al.*, "Balance and Gait Represent Independent Domains of Mobility in Parkinson Disease," *Physical Therapy*, vol. 96, no. 9, pp. 1364–1371, Sep. 2016.
- [23] J. Prince *et al.*, "Big data in parkinson's disease: Using smartphones to remotely detect longitudinal disease phenotypes," *Physiological Measurement*, vol. 39, 03 2018.
- [24] M. Djurić-Jovičić *et al.*, "Implementation of continuous wavelet transform

- mation in repetitive finger tapping analysis for patients with pd,” in *2014 22nd Telecommunications Forum Telfor (TELFOR)*, 2014, pp. 541–544.
- [25] M. Yokoe *et al.*, “Opening velocity, a novel parameter, for finger tapping test in patients with parkinson’s disease,” *Parkinsonism and Related Disorders*, vol. 15, no. 6, pp. 440–444, 2009.
- [26] X. Jia *et al.*, “Objective quantification of upper extremity motor functions in unified parkinson’s disease rating scale test,” in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 5345–5348.
- [27] C. Lee *et al.*, “A validation study of a smartphone-based finger tapping application for quantitative assessment of bradykinesia in parkinson’s disease,” *PloS one*, vol. 11, p. e0158852, 07 2016.
- [28] M. Memedi *et al.*, “Validity and responsiveness of at-home touch screen assessments in advanced parkinson’s disease,” *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 6, pp. 1829–1834, 2015.
- [29] S. Aghanavasi *et al.*, “A smartphone-based system to quantify dexterity in parkinson’s disease patients,” *Informatics in Medicine Unlocked*, vol. 9, pp. 11–17, 2017.
- [30] A. Zhan *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, 07 2018.
- [31] Ákos Jobbágy *et al.*, “Analysis of finger-tapping movement,” *Journal of Neuroscience Methods*, vol. 141, no. 1, pp. 29–39, 2005.
- [32] T. Khan *et al.*, “A computer vision framework for finger-tapping evaluation in parkinson’s disease,” *Artificial Intelligence in Medicine*, vol. 60, no. 1, pp. 27–40, 2014.
- [33] Y. Liu *et al.*, “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.
- [34] H. Li *et al.*, “Automated assessment of parkinsonian finger-tapping tests through a vision-based fine-grained classification model,” *Neurocomputing*, vol. 441, pp. 260–271, 2021.
- [35] M. R. M. Rodrigues *et al.*, “Does the Finger-to-Nose Test measure upper limb coordination in chronic stroke?” *Journal of NeuroEngineering and Rehabilitation*, vol. 14, Jan. 2017.
- [36] B. Oubre *et al.*, “Decomposition of Reaching Movements Enables Detection and Measurement of Ataxia,” *Cerebellum (London, England)*, Mar. 2021.
- [37] K. Z. Gajos *et al.*, “Computer mouse use captures ataxia and parkinsonism, enabling accurate measurement and detection,” *Movement Disorders*, vol. 35, no. 2, pp. 354–358, 2020.
- [38] D. Simonsen *et al.*, “Design and test of a microsoft kinect-based system for delivering adaptive visual feedback to stroke patients during training of upper limb movement,” *Medical & Biological Engineering & Computing*, vol. 55, no. 11, p. 1927–1935, 2017.
- [39] T. Hoffmann *et al.*, “Remote measurement via the internet of upper limb range of motion in people who have had a stroke,” *Journal of Telemedicine and Telecare*, vol. 13, no. 8, pp. 401–405, 2007.
- [40] S. Allin *et al.*, “Robust Tracking of the Upper Limb for Functional Stroke Assessment,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, p. 9, 2010.
- [41] N. Kour *et al.*, “Computer-Vision Based Diagnosis of Parkinson’s Disease via Gait: A Survey,” *IEEE Access*, vol. 7, pp. 156 620–156 645, 2019, conference Name: IEEE Access.
- [42] M. R. Ortells J, Herrero-Ezquerro MT, “Vision-based gait impairment analysis for aided diagnosis,” *Medical and Biological Engineering and Computing*, vol. 56, no. 9, pp. 1553–1564, 2018, publisher: Springer Verlag 1553.
- [43] M. Nieto-Hidalgo *et al.*, “Gait Analysis Using Computer Vision Based on Cloud Platform and Mobile Device,” Jan. 2018, iSSN: 1574-017X Pages: e7381264 Publisher: Hindawi Volume: 2018.
- [44] W. Zhu *et al.*, “A Computer Vision-Based System for Stride Length Estimation using a Mobile Phone Camera,” in *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. Reno Nevada USA: ACM, Oct. 2016, pp. 121–130.
- [45] M. Nieto-Hidalgo *et al.*, “A vision based proposal for classification of normal and abnormal gait using rgb camera,” *Journal of Biomedical Informatics*, vol. 63, pp. 82–89, 2016.
- [46] A. Toshev *et al.*, “DeepPose: Human pose estimation via deep neural networks,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 12 2013.
- [47] K. Sun *et al.*, “Deep high-resolution representation learning for human pose estimation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [48] J. Martinez *et al.*, “A simple yet effective baseline for 3d human pose estimation,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [49] Y. Rong *et al.*, “FrankMocap: Fast Monocular 3D Hand and Body Motion Capture by Regression and Integration,” *arXiv e-prints*, p. arXiv:2008.08324, Aug. 2020.
- [50] C. Zimmermann *et al.*, “3d human pose estimation in rgbd images for robotic task learning,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 1986–1992.
- [51] R. Clark *et al.*, “Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and kinect alternatives,” *Gait & Posture*, vol. 68, 11 2018.
- [52] S. Springer *et al.*, “Validity of the kinect for gait assessment: A focused review,” *Sensors*, vol. 16, no. 2, 2016.
- [53] E. Dolatabadi *et al.*, “Automated classification of pathological gait after stroke using ubiquitous sensing technology,” in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 6150–6153.
- [54] J. Andre *et al.*, “Markerless gait analysis vision system for real-time gait monitoring,” *2020 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), Autonomous Robot Systems and Competitions (ICARSC)*, 2020 IEEE International Conference on, pp. 269 – 274, 2020.
- [55] K. Sato *et al.*, “Quantifying normal and parkinsonian gait features from home movies: Practical application of a deep learning-based 2D pose estimator,” *PLOS ONE*, vol. 14, no. 11, p. e0223549, Nov. 2019.
- [56] K. Hu *et al.*, “Vision-based freezing of gait detection with anatomic directed graph representation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 4, pp. 1215–1225, 2020.
- [57] L. Kidziński *et al.*, “Deep neural networks enable quantitative movement analysis using single-camera videos,” *Nature Communications*, vol. 11, no. 1, p. 4054, Dec. 2020.
- [58] M. Capecchi *et al.*, “The kimore dataset: Kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp. 1436–1448, 2019.
- [59] T. Simon *et al.*, “Hand keypoint detection in single images using multiview bootstrapping,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [60] C. Ionescu *et al.*, “Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1325–1339, jul 2014.
- [61] L. Sigal *et al.*, “HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion,” *International Journal of Computer Vision*, vol. 87, no. 1, p. 4, 2009.
- [62] A. Savitzky *et al.*, “Smoothing and differentiation of data by simplified least squares procedures,” *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
- [63] R. Krishna *et al.*, “Quantitative assessment of cerebellar ataxia, through automated limb functional tests,” *Journal of NeuroEngineering and Rehabilitation*, vol. 16, no. 1, p. 31, Feb. 2019.
- [64] A. Al-Jawad *et al.*, “Using multi-dimensional dynamic time warping for tug test instrumentation with inertial sensors,” in *2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2012, pp. 212–218.
- [65] A. Muro-de-la Herran *et al.*, “Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications,” *Sensors*, vol. 14, no. 2, pp. 3362–3394, 2014.
- [66] J. H. Friedman, “Greedy function approximation: a gradient boosting machine,” *Annals of statistics*, pp. 1189–1232, 2001.
- [67] T. Chen *et al.*, “XGBoost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’16. New York, NY, USA: ACM, 2016, pp. 785–794.
- [68] L. M. Babrak *et al.*, “Traditional and digital biomarkers: Two worlds apart?” *Digital biomarkers*, vol. 3, no. 2, pp. 92–102, 08 2019.
- [69] H. Nagasaki, “Asymmetric velocity and acceleration profiles of human arm movements,” *Experimental Brain Research*, vol. 74, pp. 319–326, 2004.
- [70] R. N. Sawyer *et al.*, “Asymmetry of forearm rolling as a sign of unilateral cerebral dysfunction,” *Neurology*, vol. 43, no. 8, pp. 1596–1598, Aug. 1993.
- [71] E. Pretegianni *et al.*, “Eye movements in parkinson’s disease and inherited parkinsonian syndromes,” *Frontiers in neurology*, vol. 8, pp. 592–592, 11 2017.
- [72] B. Jin *et al.*, “Diagnosing parkinson disease through facial expression recognition: Video analysis,” *Journal of medical Internet research*, vol. 22, no. 7, pp. e18 697–e18 697, 07 2020.

- [73] M. R. Ali *et al.*, “Facial expressions can detect parkinson’s disease: preliminary evidence from videos collected online,” *npj Digital Medicine*, vol. 4, no. 1, p. 129, 2021.
- [74] M. Fabbri *et al.*, “Speech and voice response to a levodopa challenge in late-stage parkinson’s disease,” *Frontiers in Neurology*, vol. 8, p. 432, 2017.