Towards a Comprehensive Solution for a Vision-Based Digitized Neurological Examination

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

***Abstract*—The ability to use digitally recorded and quantified neurological exam information is important to help healthcare systems deliver better care, in-person and via telehealth, as they compensate for a growing shortage of neurologists. Current neurological digital biomarker pipelines, however, are narrowed down to a specific neurological exam component or applied for assessing specific conditions. In this paper, we propose an accessible vision-based exam and documentation solution called Digitized Neurological Examination (DNE) to expand exam biomarker recording options and clinical applications using a smartphone/tablet. Through our DNE software, healthcare providers in clinical settings 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 tests. The DNE extracts from the recorded examinations the 2D/3D human-body pose and quantifies kinematic and spatio-temporal features. The features are clinically relevant and allow clinicians to document and observe the quantified movements and the changes of these metrics over time. A web server and a user interface for recordings viewing and feature visualizations are available. DNE was evaluated on a collected dataset of 21 subjects containing normal and simulated-impaired movements. The overall accuracy of DNE is demonstrated by classifying the recorded movements using various machine learning models. Our tests show an accuracy beyond 90% for upper-limb tests and 80% for the stand-up and walk tests.**

***Index Terms*—Digital biomarkers, digitized exams, teleneurology, quantitative analysis, disease documentation, monitoring, finger tapping, finger to finger, forearm roll, stand-up 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] through teleneurology (TN) platforms. Raplidly developing TN has shown potential in making efficient assessments remotely [5]–[7] and helping in distributing scarce healthcare resources and enhancing accessibility to neurological care [8], [9]. In addition, digital biomarker exam solutions with quantification of physical evaluations that bypass clinician availability and subjectivity of assessments [10] are important to improve care and compensate for the shortage of neurologists.

Current digital biomarker exam systems are devoted to a single neurological test [11]–[13], require advanced setups/equipment [14], or lack automated assessments [15], [16]. Therefore, a digital biomarker 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 exam and documentation platform named Digitized Neurological Examination (DNE). As part of DNE, we designed an easy-to-use smartphone/tablet software with predefined 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-up and walk (SAW). These recordings are uploaded to a 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 deep-learning-based solutions such as OpenPose [17], and

Manuscript received 25 November 2021; revised 23 March 2022; accepted 11 April 2022. Date of publication 19 April 2022; date of current version 9 August 2022. This work was supported by the Jump ARCHES Endowment through the Health Care Engineering Systems Center, in part by the National Institute of Health (NIH) under Grant R01 AI139401, and in part by the Vingroup Innovation Foundation under Grant VINIF.2021.DA00128. *(Trung-Hieu Hoang and Mona Zehni contributed equally to this work.) (Corresponding authors: Trung-Hieu Hoang; Mona Zehni.)*

Trung-Hieu Hoang and Mona Zehni are with the Department of Electrical & Computer Engineering, Coordinated Science Laboratory at University of Illinois at Urbana-Champaign (UIUC), Champaign, IL 61801 USA (e-mail: hthieu@illinois.edu; mzehni2@illinois.edu).

Minh N. Do is with the Department of Electrical & Computer Engineering, Coordinated Science Laboratory at University of Illinois at Urbana-Champaign (UIUC), Champaign, IL 61801 USA, and also with the VinUni-Illinois Smart Health Center, VinUniversity, Hanoi, Vietnam (e-mail: minhdo@illinois.edu).

Huaijin Xu is with the Department of Kinesiology & Community Health at UIUC, Urbana, IL 61801 USA (e-mail: huaijin3@illinois.edu).

George Heintz is with the Healthcare Engineering Systems Center at UIUC, Urbana, IL 61801 USA (e-mail: jheintz@illinois.edu).

Christopher Zallek is with OSF HealthCare Illinois Neurological Institute–Neurology, Peoria, IL 61603 USA (e-mail: christopher.m.zallek@ini.org).

Digital Object Identifier 10.1109/JBHI.2022.3167927

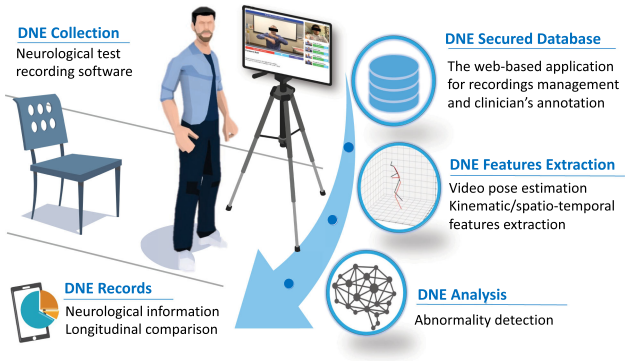


Fig. 1. Illustration of our digitized neurological exam system.

VideoPose3D [18]. From the estimated pose, unified digital biomarkers, including spatio-temporal and kinematic features, are computed [19]. We showcase the performance of our system on a dataset collected from 21 healthy subjects taking different neurological 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 illustrates the capabilities our DNE system.

We summarize the key contributions of this work as:

- We develop 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 implement a web-based dashboard for viewing the recordings and feature visualization.
- We propose a vision-based approach to study various neurological tests (FT, FTF, FR, and SAW). For each test, we define clinically interpretable kinematic and spatio-temporal quantified features.
- 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) and code will be available at https://dneproject.web.illinois.edu/.

The organization of this paper is as follows. Section II summarizes recent studies on digital biomarker 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 VIII draws our main conclusions.

II. Related Work

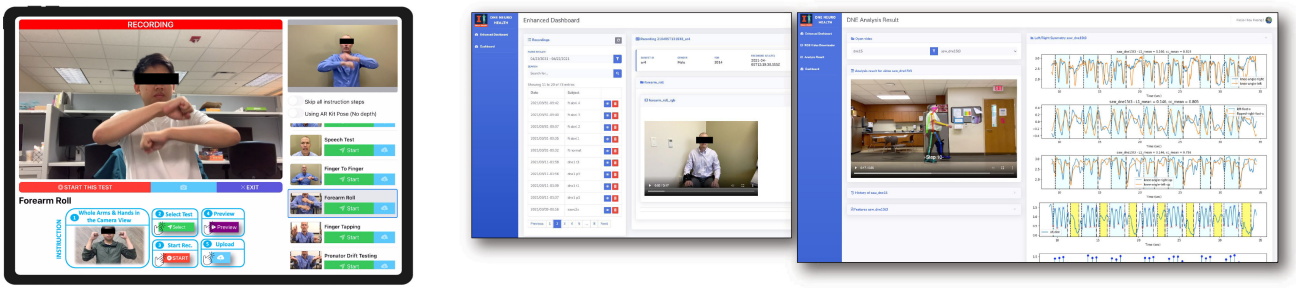
In this section, we review the related literature to different tests (FT, FTF, FR, SAW). For each test, we briefly discuss the existing sensor, web/smartphone and vision-based solutions.

Finger Tapping (FT): Sensor-based FT assessments study spectral analysis of gyroscope data [20], opening finger tap velocity captured by accelerometers [21], standard deviation, range and entropy measured by a collection of sensors including synchronized wrist watches, pressure sensors and accelerometers [22]. Several smartphone based applications [23]–[26] are designed to quantitatively evaluate various symptoms and motor skills in patients with Parkinson's 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, [11], [27]–[29] 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 [30]. While these solutions provide high accuracy, unlike our DNE, 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 [31] investigates the coordination ability of patients with chronic stroke versus healthy control using a complex marker-based motion analysis system. Oubre *et al.* [32] 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 [33]. Jaroensri *et al.* [12] 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 using 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 [14] 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 [34]. The range of motion is analyzed using an internet-based goniometer in [35]. In [36], the authors describe a vision-based system that captures upper limb motions via multiple cameras installed at different views. While this multi-camera system is less sensitive to occlusions and dynamic backgrounds, unlike our DNE system, it requires a special setup which is hard to install for home-use.

Stand-up and Walk (SAW): In our review of gait analysis literature, we focus on the marker-less [37] 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 [38]–[41]. 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 [42]–[44] brought along popular deep learning frameworks which accurately estimate the 2D/3D location of body joints from different inputs including RGB image, video and depth maps [17], [18], [45], [46]. Depth-map based gait assessment solutions relying on the estimated pose from either depth or RGBD [47], [48], have studied the rotational angle and angle velocity of certain body



(a) DNE Recorder (b) DNE Viewer

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 (best viewed in magnification).

keypoints [49] and evaluated the spatio-temporal gait metrics such as step length and time [13], [50].

Wei *et al.* [16] introduced an automated smart-phone based video capturing system with hand/body pose estimation. While neurological exams such as gait are considered in [16], feature extraction and analysis is not studied and the main focus is on the quality control of the video acquisition process. Using the estimated pose from OpenPose [17], Xue *et al.* [13] studied the remote monitoring of gait parameters for senior care. Furthermore, [51] reports timings of different segments of the timed-up-and-go (TUG) test by performing frame-based activity classification based on 2D pose data. To assess the freezing of gait (FoG) symptom in Parkinson patients, [52] proposed the use of frequency analysis methods while [53] adopted graph convolutional neural networks to attain the probability of FoG from pose data. Kidziński in [54] 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 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 tests. DNE Recorder is an iOS mobile application. It includes detailed instructions on how to perform each test alongside automated video capturing functions. Our software facilitates recording of high quality depth maps on devices equipped with LiDAR. We collect 1080×720 high-quality RGB, depth videos (upon applicable hardware) and camera calibration parameters at 60 frames per second (FPS). All recordings are synchronized into a secure cloud storage for offline processing. The user interface of this module is shown in Fig. 2(a).

DNE Analyzer: We analyze the RGB recordings offline in a separate module. The main components of DNE Analyzer include 1) vision-based pose estimation, 2) feature extraction, 3) abnormality detection.

DNE Viewer: We provide a secure web application for clinicians, neurologists and researchers to monitor raw recordings and view the analysis results from all subjects remotely. Fig. 2 (b) displays a screenshot of the DNE Viewer user interface.

IV. Dataset Collection

Our dataset collection protocol is IRB approved (#IRB.1452500) on 02/27/2020 by the University of Illinois College of Medicine at Peoria Institute Review Board 1. In this study, 21 healthy volunteers (18 females/3 males) were recruited by sampling of convenience at the OSF HealthCare Illinois Neurological Institute Outpatient Neurology Clinic (Peoria, IL). Neurological examinations examine fine motor and mobility abilities. We study the FT, FTF, FR for fine motor tasks, and evaluate the mobility by the SAW test. Below we describe in detail how these tasks are performed.

- **FT:** Participants are instructed to put their 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.
- **FR:** 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.
- **FTF:** Participants repetitively first point their index fingers towards the ceiling and then touch their fingers together in front of their chests for a duration of 15 seconds.
- **SAW:** 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 test 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, 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 11 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, we record both

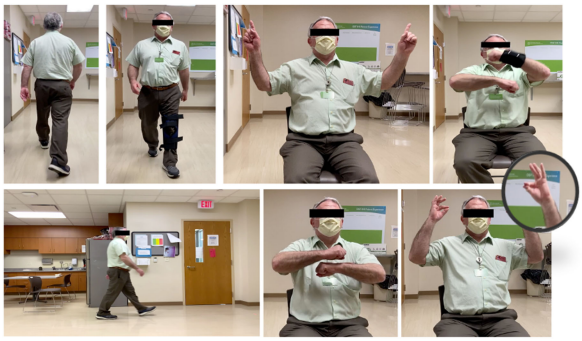


Fig. 3. Examples of DNE dataset recordings. Impairments are induced by wearing a wrist brace for FR, a rubber band for FT and a knee brace for SAW tests.

TABLE I
SUMMARY OF OUR DNE DATASET

Test	Total	Label		View		Video	
		Normal	Abnormal	Front	Side	RGB/D	RGB
FT	95	41	54	95	-	45	50
FR	92	47	45	92	-	40	52
FTF	85	41	44	85	-	45	40
SAW	103	41	62	61	42	54	49

saggital and frontal views, using iPad and iPhone, respectively. In total, including all four tests (FR, FT, FTF, SAW), we collect 375 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 [13], [38], [39], [52], [54]. The closest to our dataset is KIMORE [55] focusing on rehabilitation exercises rather than neurological tests. The 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. 3) Our dataset contains multiple data modalities, including depth videos, camera parameters, and 2D/3D pose estimation.

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 (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 various tests. We carefully design these features for each test separately to accurately reflect the subjects' performance and dedicated abnormalities. Lastly, the application layer contains several downstream tasks consuming

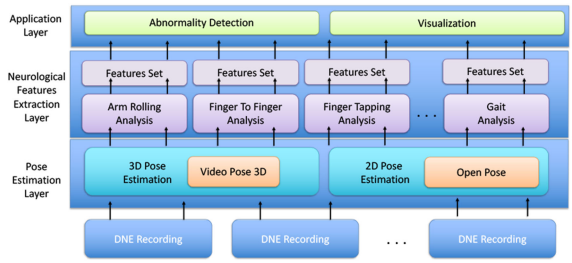


Fig. 4. Overview of DNE vision-based analysis framework.

the features, including abnormality detection and visualization for a qualitative comparison among recordings.

A. Pose Estimation

For upper body tests (FT, FTF, and FR), we use OpenPose (OP) to estimate the 2D hand [56] and body [17] pose. On the other hand, for SAW tests, we compute the 3D pose using the VideoPose3D (VP3D) package [18]. Given an RGB image, OP first detects all visible body parts and associates them to each individual by solving a graph matching problem. Meanwhile, VP3D adopts dilated temporal convolution to estimate 3D pose from sequence of 2D keypoints extracted from the video.

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 2D pose is chosen for upper body tests. However, this might not hold for the SAW test (especially depending on 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. To account for variable distance of the subjects from the camera, we normalize the estimated pose by a reference length. For FT, FTF and FR tests, the reference is the length of the forearm. For SAW, the reference is the distance between the pelvis and neck joints. We compute the reference lengths as the median of the value across all the frames. In addition, as the estimated pose can be erroneous at some frames we use median and Savitzky-Golay filtering [58]. In our dataset, we have excluded 27 recordings due to unreliable and noisy estimated pose. Therefore, we only analyzed 348 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 $\boldsymbol{v} = [v_1, \ldots, v_N]$ denote the set of N frames ordered chronologically in video $\boldsymbol{v}$. There is a one-to-one correspondence between the time associated with each frame and the frame index, where $\boldsymbol{t} = [t_1, \ldots, t_N]$ and $t_i = i/fps$, fps denoting the frame per second rate of the video. Given $\boldsymbol{v}$ and 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

TABLE II
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. [13]	✗	-	-	RGB	✗	-	-	2D	✗
Sato et al. [52]	✗	✗	✓	RGB	✗	2	2	2D	✗
Ortells et al. [38]	✓	✗	✓	Binary	✗	10	20	✗	✓
Nieto-Hidalgo et al. [39]	✓	✓	✓	Binary	✓	-	73	✗	✓
Kidzinski et al. [54]	✓	✓	✗	RGB	✗	1026	1792	2D	✗
Ours	✓	✓	✓	RGB/D	✓	21	336	2D/3D	✓

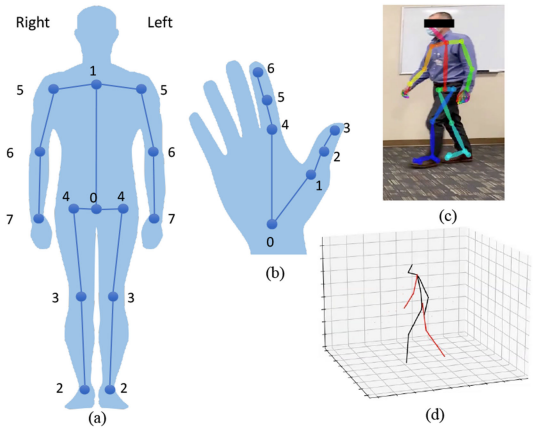


Fig. 5. Skeleton tree for (a) body B, and (b) hand H. Examples of human pose estimation (c) in 2D (B_2, H_2) using OpenPose [17] and (d) in 3D (B_3) using VideoPose3D [18].

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 $\boldsymbol{s}_{k,*}[i]$ for the location of the k-th keypoint at frame i, corresponding to skeleton tree $* \in \{H_2, B_2, B_3\}$. For $* \in \{H_2, B_2\}$, $\boldsymbol{s}_{k,*}[i] \in \mathbb{R}^2$ and for $* = B_3$, $\boldsymbol{s}_{k,*}[i] \in \mathbb{R}^3$. Furthermore, we add superscript r and l to point to right and left (R/L) body parts, respectively. For example, $\boldsymbol{s}^r_{3,H_2}[i]$ locates 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 two hands while in FR, we closely track the wrist joints.

In different tests, the subjects are asked to move certain limbs repeatedly. Thus, it is natural to compute features such as frequency, and amplitude for periodic pose patterns and report the mean and standard deviation (STD) across different cycles. In addition, for a test performed normally, the features corresponding to the R/L 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:

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

Another useful metric in our analysis is Pearson correlation coefficient denoted by CC. For two 1D discrete time series $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, we define CC as:

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

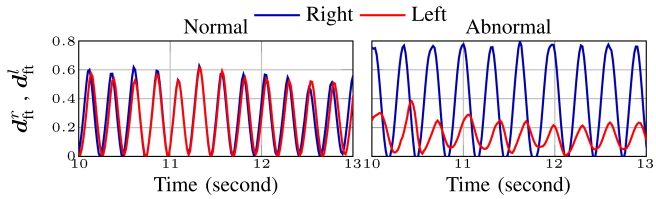


Fig. 6. FT amplitude for normal and abnormal examples.

where $\bar{.}$ and $.^T$ denote the mean and transpose operators. For highly correlated series, |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 examine the distance between the tip of the index and thumb fingers across time defined as:

$$\boldsymbol{d}^*_{\text{ft}}[i] = \|\boldsymbol{s}^*_{3,H_2}[i] - \boldsymbol{s}^*_{6,H_2}[i]\|_2, \quad * \in \{r, l\}. \tag{3}$$

Examples of $\boldsymbol{d}^r_{\text{ft}}$ and $\boldsymbol{d}^l_{\text{ft}}$ for normal and abnormal executions of the FT test are provided in Fig. 6. 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 in Fig. 6, this limits the tapping amplitude of the hand wearing the band and slows down the tapping rate. Given $\boldsymbol{d}^*_{\text{ft}}$, we compute the period for the $*$ hand, T^*_{ft}, as the time (in seconds) between two consecutive local minima (or maxima) of $\boldsymbol{d}^*_{\text{ft}}$. Frequency F^*_{ft} is the reciprocal of T^*_{ft}. We also report the magnitude of finger-tapping A^*_{ft} as the difference in consecutive minima and maxima of $\boldsymbol{d}^*_{\text{ft}}$. We also report the asymmetry of the periods (Asym($T^r_{\text{ft}}, T^l_{\text{ft}}$)), frequencies (Asym($F^r_{\text{ft}}, F^l_{\text{ft}}$)) and amplitudes (Asym($A^r_{\text{ft}}, A^l_{\text{ft}}$)) 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 $\boldsymbol{d}^r_{\text{ft}}$ and $\boldsymbol{d}^l_{\text{ft}}$ with respect to time. We adopt mean and maximum of instant speed and acceleration across tapping cycles as features. We also introduce average tapping rate as the average number of finger taps per second.

Finally, 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 between ($\boldsymbol{s}^r_{7,B_2}, \boldsymbol{s}^l_{7,B_2}$)

TABLE III
SUMMARY OF OUR DNE FEATURES

Finger Tapping (FT)	Finger to Finger (FTF)	Forearm Roll (FR)	Stand and Walk (SAW)
Amplitude (R/L) *Mean, STD, Median, Asymmetry* Maximum distance between the tip of the index and thumb fingers **Period (R/L)** *Mean, STD, Median, Asymmetry* Time (in seconds) taken to complete one tapping cycle for R/L hands **Frequency (R/L)** *Mean, STD, Median* Reciprocal of period (1/second) for R/L hands **Maximum speed (R/L)** *Mean, Asymmetry* 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 **Maximum acceleration (R/L)** *Mean, STD, Median, Asymmetry* 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 tapping rate (R/L)** Total number of finger taps divided by the duration of FT test in seconds for R/L hands **Wrist stability** *Mean, STD, Median* Variations in R/L wrist joint positions	**Horizontal symmetry** *CC* The CC of horizontal spatial trajectory of the R/L index finger **Vertical symmetry** *CC* The CC of the vertical spatial trajectory of the R/L index finger **Period (R/L)** *Mean, STD* Total time (in seconds) taken for one complete cycle (moving from the highest to the lowest vertical position and back) on each side **Average speed R/L** *Mean, STD* The traversed distance of R/L index fingers within half a cycle's period divided by half the cycle's period **Path smoothness (R/L)** *Mean, STD* The ratio between the actual traversed distance of R/L index fingers and the length of the fitted smooth curve **Velocity angle symmetry (R/L)** *Mean, STD* The pairwise CC between the angle velocity series of any two cycles	**Amplitude R/L** *Mean, STD, Median, Asymmetry* Distance between the minimum and maximimum of the vertical position of the R/L wrists **Period R/L** *Mean, STD, Median, Asymmetry* Time (in seconds) taken to complete one forearm roll cycle for R/L hands **Maximum speed (R/L)** *Mean, STD, Median, Asymmetry* 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 **Maximum acceleration (R/L)** *Mean, STD, Median, Asymmetry* 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 **Rolling speed R/L** *Mean, STD, Median* Average forearm roll speed (defined as the amplitude divided by half the rolling cycle period) **Average rolling rate R/L** Total number of forearm roll cycles divided by the duration of FR test in seconds for R/L hands	**Knee angle symmetry** *Mean, STD, Median* The CC of the aligned R/L knee angle series within a walking segment (a full pass of the room length) **Step symmetry** *Mean, STD, Median* The cycle-wise CC of the aligned spatial trajectory of the R/L foot in the horizontal axis **Step length** *Mean, STD, Median* The furthest distance between two feet within each step **Step width** *Mean, STD, Median* The shortest distance between two feet within each step **Step time** *Mean, STD, Median* The time (in seconds) to complete one step (the interval between two consecutive time-points having the shortest distance between two feet) **Time to stand** Total time taken (in seconds) from the first stand up effort to a full standing on feet state **Turning time** *Mean, STD, Median* Total time taken (in seconds) for a subject to turn around after each walking segment **Walking speed** *Mean, STD* Total of traveled distance of the pelvis joint divided by the duration of a walking segment **Cadence** *Mean, STD* Total number of steps divided by the duration of a walking segment

Asymmetry Between R/L Features is Computed Based on (1).

and $(\boldsymbol{s}^r_{6,B_2}, \boldsymbol{s}^l_{6,B_2})$ across N frames:

$$C^{\text{wrist}}_{\text{ft}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\|\boldsymbol{s}^r_{7,B_2}[i] - \boldsymbol{s}^l_{7,B_2}[i]\|_2}{\|\boldsymbol{s}^r_{7,\text{B}_2}[i]\|_2}, \tag{4}$$

$$C^{\text{elbow}}_{\text{ft}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\|\boldsymbol{s}^r_{6,B_2}[i] - \boldsymbol{s}^l_{6,B_2}[i]\|_2}{\|\boldsymbol{s}^r_{6,B_2}[i]\|_2}. \tag{5}$$

Finger to Finger (FTF): In 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 than the outer fingertip. Hence, we focus on this joint for FTF test. In a normal FTF, the horizontal and vertical trajectories of the R/L hands are symmetric up to a mirroring (Fig. 7(a) top row), while this does not necessarily hold for abnormal case (Fig. 7(a) bottom row). Thus, in each cycle, we define the cross correlation of the R/L horizontal (x) and vertical (y) coordinates as the horizontal ($S^{\text{finger-x}}_{\text{ftf}}$) and vertical symmetries ($S^{\text{finger-y}}_{\text{ftf}}$):

$$S^{\text{finger-x}}_{\text{ftf}} = \text{CC}\left(\left[\boldsymbol{s}^l_{5,H_2}\right]_x, -\left[\boldsymbol{s}^r_{5,H_2}\right]_x\right), \tag{6}$$

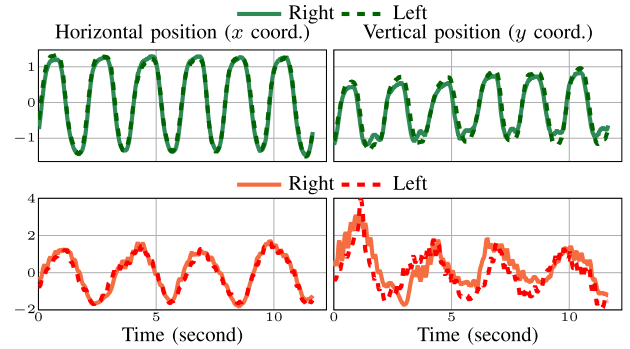
$$S^{\text{finger-y}}_{\text{ftf}} = \text{CC}\left(\left[\boldsymbol{s}^l_{5,H_2}\right]_y, \left[\boldsymbol{s}^r_{5,H_2}\right]_y\right) \tag{7}$$

where $[\boldsymbol{s}^*_{5,H_2}]_\dagger = \{\boldsymbol{s}^*_{5,H_2}[i](\dagger)\}_{i=1}^N$, $\dagger \in \{x, y\}$ and $* \in \{r, l\}$, 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.

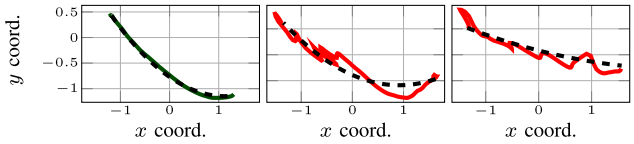
Patients with neurological impairments tend to have tremors while moving their fingers during FTF test [59]. This leads to a deviation of the fingers' trajectory from a smooth curve. To characterize this deviation, we first fit a smooth curve to the fingers' trajectory, in the form of a second order polynomial in terms of the x and y coordinates. 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 R/L hands. Examples of normal and abnormal finger trajectories alongside the smooth fitted curves are plotted in Fig. 7(b).

Another feature we found helpful in detecting abnormal function in FTF is instant velocity. We derive the instant velocity vector by the first derivative of the horizontal and vertical pose with respect to time. We then examine the angle between the vertical and horizontal components of this vector on the R/L hands. At time instant t, the *velocity angle* $\boldsymbol{\theta}$ is:

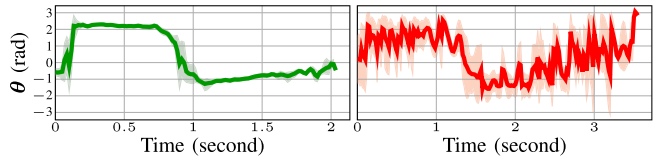
$$\boldsymbol{\theta}^*(t) = \text{atan2}\left(\frac{\frac{d\left[\boldsymbol{s}^*_{5,H_2}\right]_y}{dt}}{\frac{d\left[\boldsymbol{s}^*_{5,H_2}\right]_x}{dt}}\right), * \in \{r, l\}. \tag{8}$$



(a) Horizontal and vertical (first/second column) locations of joint index 5 in H_2 for normal and abnormal (first/second row) recordings.



(b) Actual trajectory (solid) vs. smoothed fitted trajectory (dashed).



(c) Cycle-wise average of velocity angle.

Fig. 7. FTF features including finger (a) positions, (b) spatial trajectory, (c) velocity angle. Green (red) curves stand for normal (abnormal) recordings. In each row (column), the subplots share the same vertical (horizontal) axis.

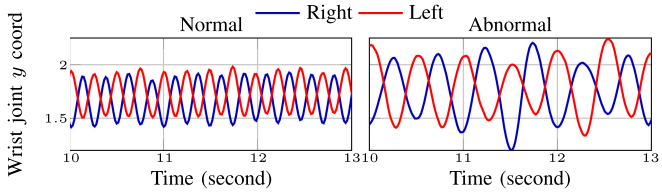
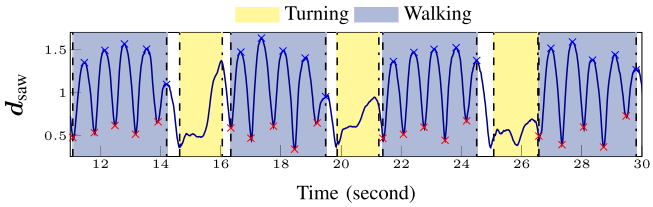


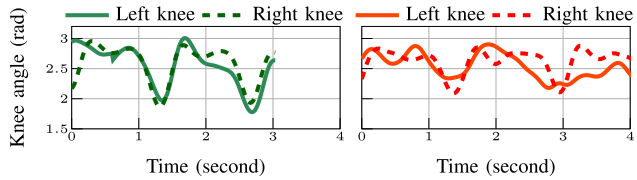
Fig. 8. Vertical (y) coordinate of the wrist joint versus time for normal and abnormal examples in FR test.

Next, for each hand, we compare $\boldsymbol{\theta}$ across different cycles using CC 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 STD. Examples of normal and abnormal aligned velocity angles across different cycles are provided in Fig. 7(c). Note that for abnormal FTF, large magnitude fluctuations, caused by tremors in moving the hands, visibly appear in $\boldsymbol{\theta}$.

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} 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,



(a) Distance between two feet (d_{saw}) versus time. Marker × and × denote the start and end 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 a walking segment taken from a normal and an abnormal recording of the same subject. Green (red) curves stand for normal (abnormal) recordings.

Fig. 9. Examples of SAW features.

we include the asymmetry of the aforementioned metrics in the FR features.

We also 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 $C^{\text{elbow}}_{\text{fr}}$ and define it analogous to (5).

Stand-up and Walk (SAW): We use the side-view SAW recordings in our analysis of SAW test. For SAW pose estimation, we use VP3D [18]. 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 $\boldsymbol{s}_{0,B_2}$ 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. Supplementary Fig. S3 visualizes these segments.

For the SU segment of SAW, we focus on the time to stand [60], 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 segment are obtained based on the distance between the two feet stated as:

$$\boldsymbol{d}_{\text{saw}}[i] = ||\boldsymbol{s}^r_{2,B_3}[i] - \boldsymbol{s}^l_{2,B_3}[i]||_2. \tag{9}$$

Note that, the periodic nature of a normal gait also reflects in $\boldsymbol{d}_{\text{saw}}$ (see Fig. 9(a)). Given $\boldsymbol{d}_{\text{saw}}$, we highlight different W and TU segments in Fig. 9(a). For a gait pattern derived based on $\boldsymbol{d}_{\text{saw}}$, *step time* is the time to complete one step and computed

as the time difference between two consecutive local maxima of $\boldsymbol{d}_{\text{saw}}$. Meanwhile, *step length* defined as linear distance between two successive placements of the same foot [61] manifests as the local maxima of $\boldsymbol{d}_{\text{saw}}$. The *step width*, on the other hand, is interpreted as the local minima of $\boldsymbol{d}_{\text{saw}}$. The calculations of these features in turning segments are excluded.

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. 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 of the R/L 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). For each frame, we define the knee angle as the angle between $\boldsymbol{s}^r_{3,B_3} - \boldsymbol{s}^r_{4,B_3}$ and $\boldsymbol{s}^r_{3,B_3} - \boldsymbol{s}^r_{2,B_3}$ for the right leg and $\boldsymbol{s}^l_{3,B_3} - \boldsymbol{s}^l_{4,B_3}$ and $\boldsymbol{s}^l_{3,B_3} - \boldsymbol{s}^l_{2,B_3}$ 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. 9(b). For normal gait, the R/L knee angles are highly correlated after alignment (Fig. 9(b) top row), while this does not hold for abnormal gait (Fig. 9(b) bottom row).

In addition, we define step symmetry between the R/L 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 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_P/N_T and A_P/A_T denote the normal and abnormal recordings captured by iPhone/iPad.

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

Fig. 10 illustrates the distribution of A-A, N-N, and N-A distances across 20 different subjects for a subset of features of FTF test. 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.

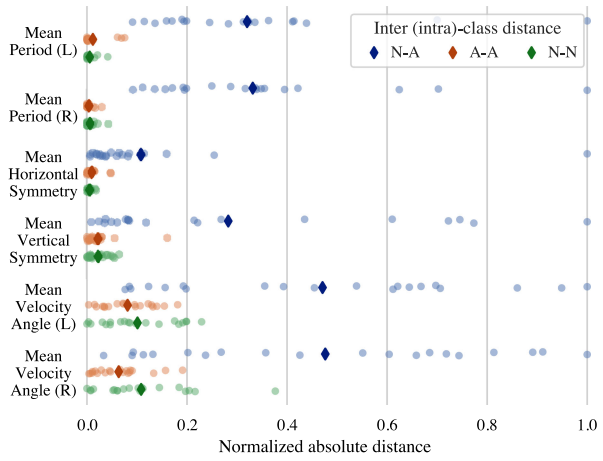


Fig. 10. The inter-class and intra-class distances between some features of normal *(N)* and abnormal *(A)* FTF recordings. ♦ denotes the mean value.

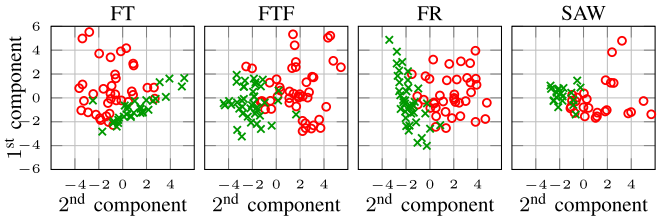


Fig. 11. PCA analysis of FT, FTF, FR, and SAW tests. Green crosses and red circles stand for normal and abnormal recordings. All subplots share the same axis.

B. Abnormality Detection

Principal component analysis (PCA): 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 PCA. For this purpose, for each test, we concatenate the set of features listed in Table III and normalize them before passing to PCA. Fig. 11 showcases the results for different tests. It is observed that the normal and abnormal recordings are separated in dimension reduced feature space. This implies that our defined features are descriptive and well differentiate normal from abnormal.

Abnormal Class Distribution: In Fig. 12 we compare the distribution of normal versus abnormal features for FT, FTF, FR, and SAW tests. These plots clearly indicate the difference in distribution between two classes. Normal features are more concentrated in a specific range, however the abnormal features are often less regular and have a higher STD.

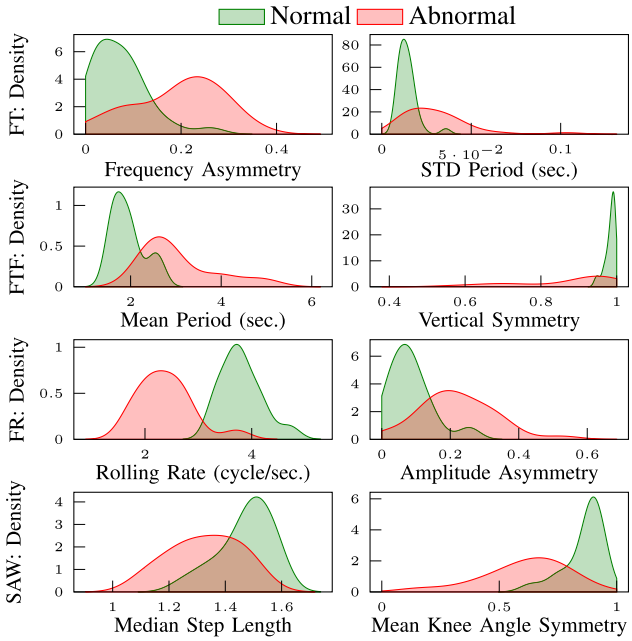


Fig. 12. Distribution of normal/abnormal features for FT, FTF, FR, SAW tests plotted in first, second, third and last rows. We used kernel density estimation to fit distributions to the data.

Abnormality Detection: We assess the normal and abnormal classification performance using our features. Therefore, we utilize several machine learning (ML) models that are grouped into: 1) tree-based methods such as Random Forest (RF), Gradient-Boosting Machine (GBM) [62], XGBoost [63] and 2) parametric models trained using gradient-descent updates, including Logistic Regression (LR), Support Vector Machine with radial basis function (RBF) kernel (RSVM) and Multi-layer Perceptron (MLP) with rectified linear unit (ReLU) activation.

We also benchmarked our ML classification performance against two deep learning (DL) baselines. Both DL models predict normal versus abnormal based on major keypoint pose sequence, unlike the ML based models which perform classification on the extracted spatio-temporal/kinematic features. In the first DL baseline, we adopt a long-short term memory (LSTM) [64] based sequential model while in the second DL approach, similar to [65], we use convolutional neural networks (CNN). Details of ML and DL based classification models, data processing and hyper parameters are provided in the Supplementary Section II and Table S1. We evaluate different models via metrics such as accuracy, average precision, F1 score, and area under the ROC curve (AUC).

We have two splitting schemes to separate the train from test sets. 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.

We perform 5-fold cross validation and summarize the average classification performance of all ML and DL models in Table IV. While all models perform well for various tests, among ML models RSVM and GBM/XGBoost tend to perform better on most metrics. However, the gap between the performance of all ML models is not significant. This suggests that the extracted set of features well-distinguish normal from abnormal samples.

Furthermore, comparing ML and DL models, we notice that: 1) While DL models perform well on FT, FR and FTF tasks (especially for video-based splitting), they are lagging behind ML models for SAW. We attribute this to the fact that SAW involves more complex motion patterns. Therefore, DL models require larger datasets to be able to learn the classification task from the pose data. 2) DL features extracted from the pose data lack clinical interpretability. 3) For subject-based splitting, ML models operating on the spatio-temporal/kinematic features outperform DL models on most metrics. This indicates better generalization capability of our features on unseen subjects compared to DL models operating on pose data.

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 RF models. This analysis gives us the weight of all features, sorted in descending order in Supplementary Fig. S4.

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 (Supplementary Fig. S4(a)) and FTF (Supplementary Fig. S4(c)), the features with the largest weights are frequency asymmetry and horizontal ($S_{\text{ftf}}^{\text{finger-x}}$) symmetry, 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 [66]–[68] 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. 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. This result also affirms our observations in Fig. 12.

VII. Discussion & Challenges

In this section, we discuss various aspects of DNE including feature design, robustness, clinical relevance and application as well as the current challenges and our proposed solutions.

A. Discussion

Feature Design: The main goal of our DNE system is to provide an objective tool for quantifying and documentation of recordings of neurological tests. Thus, it is critical to design a set of clinical interpretable features that explain the performance

TABLE IV

CLASSIFICATION PERFORMANCE OF SEVERAL MACHINE LEARNING MODELS, INCLUDING RANDOM FOREST (RF), GRADIENT-BOOSTING MACHINE (GBM), XGBOOST, LOGISTIC REGRESSION (LR), SUPPORT VECTOR MACHINE WITH RBF KERNEL (RSVM), AND MULTI-LAYER PERCEPTRON (MLP) ALONGSIDE LSTM AND CNN BASED DEEP LEARNING MODELS FOR FT, FTF, FR AND SAW TESTS

		Subject Based							Video Based						
Test	**Model**	**Acc**	**Precision**	**Recall**	**Specificity**	**F1 Score**	**AUC**	**AP**	**Acc**	**Precision**	**Recall**	**Specificity**	**F1 Score**	**AUC**	**AP**
FT	RF	0.8554	0.8947	0.8500	0.8750	0.8500	0.8625	0.8339	0.8773	0.9278	0.8492	0.9236	0.8672	0.8864	0.8643
	GBM	**0.8804**	0.9156	**0.8750**	0.9000	**0.8742**	**0.8875**	0.8602	0.8866	0.9464	0.8470	**0.9418**	0.8839	0.8944	0.8864
	XGBOOST	0.8304	0.8778	0.8250	0.8500	0.8263	0.8375	0.8049	0.8655	0.9206	0.8292	0.9218	0.8481	0.8755	0.8530
	LR	0.8679	**0.9714**	0.7750	**0.9750**	0.8514	0.8750	**0.8732**	0.8773	0.9492	0.8292	**0.9418**	0.8613	0.8855	0.8775
	RSVM	0.8679	0.8950	**0.8750**	0.8750	0.8639	0.8750	0.8428	0.8916	0.9014	**0.8914**	0.8951	0.8867	0.8932	0.8631
	MLP	0.8679	0.9350	0.8250	0.9250	0.8575	0.8750	0.8578	0.8563	0.9300	0.8029	0.9236	0.8199	0.8632	0.8387
	LSTM	0.8089	0.8273	0.8250	0.8000	0.8146	0.8125	0.7705	**0.9008**	0.9492	0.8796	**0.9418**	**0.9029**	**0.9107**	**0.9044**
	CNN	0.8304	0.8273	**0.8750**	0.7833	0.8474	0.8292	0.8024	0.8916	**0.9514**	0.8514	**0.9418**	0.8730	0.8966	0.8852
FTF	RF	0.8625	0.9232	0.8250	0.9000	0.8510	0.8625	0.8357	0.9623	0.9550	0.9818	0.9400	0.9666	0.9609	0.9473
	GBM	0.9125	0.9378	0.9000	0.9250	0.8993	0.9125	0.8878	**0.9895**	0.9800	**1.0000**	0.9800	**0.9895**	**0.9900**	**0.9800**
	XGBOOST	**0.9250**	0.9278	**0.9250**	0.9250	**0.9249**	**0.9250**	**0.9028**	0.9684	0.9800	0.9636	0.9800	0.9704	0.9718	0.9647
	LR	0.8375	0.9378	0.7500	0.9250	0.8004	0.8375	0.8128	0.8930	0.9314	0.8805	0.9200	0.8988	0.9003	0.8853
	RSVM	0.8875	0.9378	0.8500	0.9250	0.8708	0.8875	0.8628	0.9579	0.9600	0.9636	0.9600	0.9599	0.9618	0.9447
	MLP	0.8625	0.8788	0.8750	0.8500	0.8619	0.8625	0.8218	0.9789	0.9778	0.9778	0.9800	0.9778	0.9789	0.9686
	LSTM	0.8875	**1.0000**	0.7750	**1.0000**	0.8338	0.8875	0.8875	0.9789	0.9800	0.9818	0.9800	0.9799	0.9809	0.9723
	CNN	0.8875	0.9492	0.8250	0.9500	0.8735	0.8875	0.8688	0.9684	**1.0000**	0.9455	**1.0000**	0.9705	0.9727	0.9770
FR	RF	0.8250	0.9100	0.7500	0.9000	0.8040	0.8250	0.7975	0.8737	0.8656	0.8583	0.8873	0.8551	0.8728	0.8093
	GBM	0.8500	0.8878	0.8250	0.8750	0.8360	0.8500	0.8128	0.9033	0.9124	0.8806	0.8936	0.8914	0.8871	0.8543
	XGBOOST	0.8500	0.9100	0.8000	0.9000	0.8325	0.8500	0.8225	0.8947	0.9064	0.8583	0.9255	0.8742	0.8919	0.8406
	LR	0.8625	**0.9500**	0.7750	**0.9500**	0.8414	0.8625	0.8500	0.8947	0.9492	0.8083	0.9618	0.8635	0.8851	0.8513
	RSVM	**0.9125**	0.9278	**0.9000**	0.9250	**0.9097**	**0.9125**	**0.8903**	0.8717	0.8850	0.8417	0.9055	0.8567	0.8736	0.8221
	MLP	0.7875	0.8955	0.7000	0.8750	0.7413	0.7875	0.7580	0.8132	0.8337	0.8000	0.7891	0.8084	0.7945	0.7605
	LSTM	0.8875	0.9100	0.8750	0.9000	0.8859	0.8875	0.8600	0.8507	0.8929	0.7667	0.9255	0.8227	0.8461	0.7959
	CNN	0.8000	0.8700	0.7250	0.8750	0.7761	0.8000	0.7650	**0.9539**	**1.0000**	**0.9222**	**1.0000**	**0.9568**	**0.9611**	**0.9683**
SAW	RF	0.7877	0.8167	0.7917	0.7946	0.7804	0.7932	0.7542	0.8000	0.8679	0.8429	0.7467	0.8385	0.7948	0.8270
	GBM	0.8189	**0.9000**	0.7917	**0.8571**	0.8042	0.8244	0.7958	0.8200	0.8406	0.9000	0.6967	0.8561	0.7983	0.8139
	XGBOOST	0.8261	0.8250	0.8542	0.8036	0.8240	0.8289	0.7677	0.8200	0.8317	**0.9333**	0.6300	0.8670	0.7817	0.8106
	LR	0.8189	0.8375	0.8542	0.7946	0.8250	0.8244	0.7802	0.7800	0.8762	0.7810	**0.7867**	0.8097	0.7838	0.8257
	RSVM	**0.8606**	0.8500	**0.9375**	0.7946	**0.8740**	**0.8661**	0.8187	**0.8400**	**0.8929**	0.8714	**0.7867**	**0.8685**	**0.8290**	**0.8514**
	MLP	0.8189	0.8500	0.8542	0.7946	0.8240	0.8244	0.7771	0.7800	0.8179	0.8714	0.6467	0.8277	0.7590	0.7860
	LSTM	0.7372	0.8333	0.6250	0.8393	0.6778	0.7321	0.6950	0.7800	0.7833	0.8648	0.6700	0.8139	0.7674	0.7541
	CNN	0.7877	0.8542	0.7292	0.8393	0.7643	0.7842	0.7452	0.7800	0.8267	0.8076	0.7500	0.8063	0.7788	0.7962

The best and second best results are in **bold** and underline, respectively.

of a subject on various motor tasks. In addition, having powerful digital biomarkers reduces the workload of normal versus abnormal classification models and improves their generalization, especially when large training datasets are not available. Furthermore, unlike black-box DL models, the explainability of our diverse set of features allows clinicians to better understand and track patients' status over time.

Robustness: DNE is resilient to changes in slight deviations from the camera view, distance to the camera, subject clothing, and mild pixel intensity changes due to intermediate data standardization and robust pose estimation steps (Section V-A and V-B). This is experimentally shown by the low intra-class feature distances in Fig. 10. Data normalization and filtering in the pre-processing step also helps in eliminating noise and propagated errors from the pose estimation module.

In FT, FR and SAW tests, the abnormality in the motion is imposed by wearing equipment which are visible in the recordings. The pose estimation models we have used (OP and VP3D) are robust to the appearance of the equipment and can accurately predict the joint locations regardless of the presence of the equipment. The features incorporated in the classification tasks are derived from the pose data. Therefore, the quantified features and the classification performance is not affected by the visual cues from the equipment.

Clinical Relevance: In our dataset, the abnormalities in the movements of the subjects were simulated. The simulated impairment in the FR test is the closest to what is observed in clinics for patients with neurological disorders. In the simulated impairment for FR, the arm with no moulage satellites around the weighted wrist, causing a decrease in the orbit frequency (Fig. 8). This is coherent with the clinical observations of patients with neurological impairments.

In the FTF test, the simulated abnormality would be more realistic, if the tremor or inaccuracy of movement increased as the finger got closer to its target (i.e. when the two fingers approach). In our current dataset, the subjects often simulated the tremor throughout their movements which is only seen in severe cases. In addition, for the FT test, often the abnormality is a combination of decreased amplitude and rate (Fig. 6) and in Parkinson's decrements of both. In our DNE dataset, some subjects simulated more of one or the other.

In SAW, the abnormality in real patients appear as a combination of slow time-to-stand, decreased step length, increased step-time, and asymmetry of gait features. In our dataset, the abnormality was imposed by wearing a knee brace. Alongside asymmetry between the R/L knee angles, we observed decreased step-length for the subjects wearing the knee brace (Fig. 12-SAW). These are in-line with clinical observations from real patients.

Overall, features that clinicians observe were disrupted from normal findings to various degrees, although the pattern of disruption of features may have not been exact for a specific condition. We showed that DNE was able to define clinically interpretable features and detect differences between normal and simulated impaired recordings. As future work, to expand its clinical impact, we will focus our analysis on real patients with various neurological impairment severity levels, and with other neurological tests, such as eye movement [69], facial activation [70], [71], or phonation [72].

Clinical Application: The initial clinical application of DNE is measuring and documenting features of various neurological exams. This would allow for improved communication of objective exam quantification and the ability to assess for changes over time. As future work, with clinicians' supervision, we will examine and report the performance of DNE on real patients. A longer term goal is to assist clinicians with classification of recordings and provide a platform for longitudinal monitoring of patients.

B. Challenges

Depth Ambiguity: Analyzing human motion from 2D RGB data requires dealing with uncertainties associated with lacking depth information. Furthermore, depth ambiguity becomes a more prominent challenge for the SAW test with frontal view recordings rather than sagittal view. It also avoids defining the spatial features in their absolute units. Currently, to mitigate the issues corresponding to these depth uncertainties, for upper limb tests, the subjects are asked to perform the tests while facing the camera and (roughly) in parallel to its image plane. In our processing steps, we also perform pose normalizations to compensate for scale variations due to variable distance from the camera. To further address this issue, we believe incorporating LiDAR depth maps captured by recent iOS devices, in the pose estimation step can prove helpful.

Self-baselining: 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 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. 10). 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.

Real-time DNE: Our current DNE system and the extracted kinematic/spatio temporal features rely on tracking the human pose from the video recordings in an offline step using off-the-shelf pose estimation modules. Currently, the pose estimation step is the most computationally expensive step, hindering real time processing and feature extraction. To address this challenge, on-device lighter pose estimation models (with small sacrifice on the accuracy), that focus on extracting major keypoints rather than the whole body pose are necessary.

VIII. Conclusion

In this paper, we proposed a comprehensive vision-based digital biomarker exam solution named Digitized 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 375 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 classification accuracy and F1 scores using a variety of machine learning models. Future work will 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 Neurol.*, vol. 78, no. 2, pp. 165–176, 2021.

[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.

[4] M. A. Hussona *et al.*, "The virtual neurologic exam: Instructional videos and guidance for the COVID-19 era," *Can. J. Neurological Sci.*, vol. 47, no. 5, pp. 598–603, May 2020.

[5] 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 Scler. Frontotemporal Degeneration*, vol. 21, no. 7-8, pp. 542–548, Nov. 2020.

[6] C. Duncan *et al.*, "Video consultations in ordinary and extraordinary times," *Practical Neurol.*, vol. 20, no. 5, pp. 396–403, Oct. 2020.

[7] V. Patterson, "Neurological telemedicine in the COVID-19 era," *Nature Rev. Neurol.*, vol. 17, no. 2, pp. 73–74, Feb. 2021.

[8] S. A. Mutgi *et al.*, "Emerging subspecialties in neurology: Telestroke and teleneurology," *Neurology*, vol. 84, no. 22, pp. e191–e193, 2015.

[9] E. R. Dorsey *et al.*, "Teleneurology and mobile technologies: The future of neurological care," *Nature Rev. Neurol.*, vol. 14, no. 5, pp. 285–297, 2018.

[10] J. L. Adams *et al.*, "Digital technology in movement disorders: Updates, applications, and challenges," *Curr. Neurol. Neurosci. Rep.*, vol. 21, no. 4, Mar. 2021, Art. no. 16.

[11] S. Williams *et al.*, "The discerning eye of computer vision: Can it measure Parkinson's finger tap bradykinesia," *J. Neurological Sci.*, vol. 416, 2020, Art. no. 117003.

[12] R. Jaroensri *et al.*, "A video-based method for automatically rating ataxia," in *Proc. 2nd Mach. Learn. Healthcare Conf.*, vol. 68, 2017, pp. 204–216.

[13] D. Xue *et al.*, "Vision-based gait analysis for senior care," *Mach. Learn. Health (ML4H) Workshop NeurIPS*, 2018.

[14] V. Tedim Cruz *et al.*, "A novel system for automatic classification of upper limb motor function after stroke: An exploratory study," *Med. Eng. Phys.*, vol. 36, no. 12, pp. 1704–1710, Dec. 2014.

[15] M. Grobe-Einsler *et al.*, "Development of SARA^{home}, a new video-based tool for the assessment of ataxia at home," *Movement Disord.*, vol. 36, no. 5, pp. 1242–1246, 2021.

[16] Y. Wei *et al.*, "Interactive video acquisition and learning system for motor assessment of Parkinson's disease," in *Proc. 13th Int. Joint Conf. Artif. Intell.*, Z.- H. Zho, Ed., 2021, pp. 5024–5027.

[17] Z. Cao *et al.*, "Openpose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, Jan. 2021.

[18] D. Pavllo *et al.*, "3D human pose estimation in video with temporal convolutions and semi-supervised training," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 7753–7762.

[19] D. A. Winter, "Biomechanics and Motor Control of Human Movement," *Kinematics*, New York, NY, USA: Wiley, 2009, ch. 3, pp. 45–81.

[20] M. Djurić-Jovičić *et al.*, "Implementation of continuous wavelet transformation in repetitive finger tapping analysis for patients with PD," in *Proc. 22nd Telecommun. Forum Telfor*, 2014, pp. 541–544.

[21] M. Yokoe *et al.*, "Opening velocity, a novel parameter, for finger tapping test in patients with Parkinson's disease," *Parkinsonism Related Disord.*, vol. 15, no. 6, pp. 440–444, 2009.

[22] X. Jia *et al.*, "Objective quantification of upper extremity motor functions in unified Parkinson's disease rating scale test," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 5345–5348.

[23] 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, Jul. 2016, Art. no. e0158852.

[24] M. Memedi *et al.*, "Validity and responsiveness of at-home touch screen assessments in advanced Parkinson's disease," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 6, pp. 1829–1834, Nov. 2015.

[25] S. Aghanavesi *et al.*, "A smartphone-based system to quantify dexterity in Parkinson's disease patients," *Informat. Med. Unlocked*, vol. 9, pp. 11–17, 2017.

[26] A. Zhan *et al.*, "Using smartphones and machine learning to quantify Parkinson disease severity: The mobile Parkinson disease score," *JAMA Neurol.*, vol. 75, no. 7, pp. 876–880, Jul. 2018.

[27] Ákos Jobbágy *et al.*, "Analysis of finger-tapping movement," *J. Neurosci. Methods*, vol. 141, no. 1, pp. 29–39, 2005.

[28] T. Khan *et al.*, "A computer vision framework for finger-tapping evaluation in Parkinson's disease," *Artif. Intell. Med.*, vol. 60, no. 1, pp. 27–40, Jan. 2014.

[29] Y. Liu *et al.*, "Vision-based method for automatic quantification of parkinsonian bradykinesia," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 1952–1961, Oct. 2019.

[30] 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.

[31] M. R. M. Rodrigues *et al.*, "Does the finger-to-nose test measure upper limb coordination in chronic stroke," *J. Neuroengineering Rehabil.*, vol. 14, no. 1, pp. 1–11, Jan. 2017.

[32] B. Oubre *et al.*, "Decomposition of reaching movements enables detection and measurement of ataxia," *Cerebellum*, vol. 20, no. 6, pp. 811–822, Mar. 2021.

[33] K. Z. Gajos *et al.*, "Computer mouse use captures ataxia and Parkinsonism, enabling accurate measurement and detection," *Movement Disord.*, vol. 35, no. 2, pp. 354–358, 2020.

[34] 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," *Med. Biol. Eng. Comput.*, vol. 55, no. 11, p. 1927–1935, 2017.

[35] T. Hoffmann *et al.*, "Remote measurement via the internet of upper limb range of motion in people who have had a stroke," *J. Telemed. Telecare*, vol. 13, no. 8, pp. 401–405, 2007.

[36] S. Allin *et al.*, "Robust tracking of the upper limb for functional stroke assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 5, pp. 542–550, Oct. 2010.

[37] 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.

[38] M. R. Ortells and J. Herrero-Ezquerro MT, "Vision-based gait impairment analysis for aided diagnosis," *Med. Biol. Eng. Comput.*, vol. 56, no. 9, pp. 1553–1564, 2018.

[39] M. Nieto-Hidalgo *et al.*, "Gait analysis using computer vision based on cloud platform and mobile device," *Mobile Info. Syst.,* vol. 2018, Jan. 2018, Art. no. 7381264.

[40] W. Zhu *et al.*, "A computer vision-based system for stride length estimation using a mobile phone camera," in *Proc. 18th Int. Conf. Comput. Accessibility*, 2016, pp. 121–130.

[41] M. Nieto-Hidalgo *et al.*, "A vision based proposal for classification of normal and abnormal gait using RGB camera," *J. Biomed. Informat.*, vol. 63, pp. 82–89, 2016.

[42] A. Toshev *et al.*, "Deeppose: Human pose estimation via deep neural networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1653–1660.

[43] K. Sun *et al.*, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5686–5696.

[44] J. Martinez *et al.*, "A simple yet effective baseline for 3D human pose estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2640–2649.

[45] Y. Rong *et al.*, "Frankmocap: A monocular 3D whole-body pose estimation system via regression and integration," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2021, pp. 1749–1759.

[46] C. Zimmermann *et al.*, "3D human pose estimation in RGBD images for robotic task learning," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 1986–1992.

[47] 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, pp. 193–200, Nov. 2018.

[48] S. Springer *et al.*, "Validity of the Kinect for gait assessment: A focused review," *Sensors*, vol. 16, no. 2, 2016, Art. no. 194.

[49] E. Dolatabadi *et al.*, "Automated classification of pathological gait after stroke using ubiquitous sensing technology," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2016, pp. 6150–6153.

[50] J. Andre *et al.*, "Markerless gait analysis vision system for real-time gait monitoring," in *Proc. IEEE Int. Conf. Auton. Robot Syst. Competitions*, 2020, pp. 269–274.

[51] T. Li *et al.*, "Automatic timed up-and-go sub-task segmentation for Parkinson's disease patients using video-based activity classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 11, pp. 2189–2199, Nov. 2018.

[52] 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, Nov. 2019, Art. no. e0223549.

[53] K. Hu *et al.*, "Vision-based freezing of gait detection with anatomic directed graph representation," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 4, pp. 1215–1225, Apr. 2020.

[54] L. Kidziński *et al.*, "Deep neural networks enable quantitative movement analysis using single-camera videos," *Nature Commun.*, vol. 11, no. 1, Dec. 2020, Art. no. 4054.

[55] M. Capecci *et al.*, "The KIMORE dataset: Kinematic assessment of MOvement and clinical scores for remote monitoring of physical rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 7, pp. 1436–1448, Jul. 2019.

[56] T. Simon *et al.*, "Hand keypoint detection in single images using multiview bootstrapping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1145–1153 .

[57] Y. Wu *et al.*, "Detectron2," 2019. [Online]. Available: https://github.com/facebookresearch/detectron2

[58] A. Savitzky *et al.*, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964.

[59] R. Krishna *et al.*, "Quantitative assessment of cerebellar ataxia, through automated limb functional tests," *J. Neuroengineering Rehabil.*, vol. 16, no. 1, Feb. 2019, Art. no. 31.

[60] A. Al-Jawad *et al.*, "Using multi-dimensional dynamic time warping for TUG test instrumentation with inertial sensors," in *Proc. IEEE Int. Conf. Multisensor Fusion Integration Intell. Syst.*, 2012, pp. 212–218.

[61] 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.

[62] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, 2001.

[63] T. Chen *et al.*, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794.

[64] S. Hochreiter *et al.*, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[65] Ł. Kidziński *et al.*, "Automatic real-time gait event detection in children using deep neural networks," *PLoS One*, vol. 14, no. 1, 2019, Art no. e0211466.

[66] L. M. Babrak *et al.*, "Traditional and digital biomarkers: Two worlds apart," *Digit. Biomarkers*, vol. 3, no. 2, pp. 92–102, Aug. 2019.

[67] H. Nagasaki, "Asymmetric velocity and acceleration profiles of human arm movements," *Exp. Brain Res.*, vol. 74, pp. 319–326, 2004.

[68] 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.

[69] E. Pretegiani *et al.*, "Eye movements in Parkinson's disease and inherited Parkinsonian syndromes," *Front. Neurol.*, vol. 8, pp. 592–592, Nov. 2017.

[70] B. Jin *et al.*, "Diagnosing Parkinson disease through facial expression recognition: Video analysis," *J. Med. Internet Res.*, vol. 22, no. 7, pp. e18 697–e18 697, Jul. 2020.

[71] M. R. Ali *et al.*, "Facial expressions can detect Parkinson's disease: Preliminary evidence from videos collected online," *NPJ Digit. Med.*, vol. 4, no. 1, 2021, Art. no. 129.

[72] M. Fabbri *et al.*, "Speech and voice response to a levodopa challenge in late-stage Parkinson's disease," *Front. Neurol.*, vol. 8, 2017, Art. no. 432.