The discerning eye of computer vision: can it measure Parkinson's finger tap bradykinesia?

DOI:
10.1016/j.jns.2020.117003

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Journal of Neurological Sciences.

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
The discerning eye of computer vision: can it measure Parkinson’s finger tap bradykinesia?

Authors: Stefan Williams, MBBS, BSc; Zhibin Zhao, BEng; Awais Hafeez, MS, BSc; David C Wong, PhD; Samuel D Relton, PhD; Hui Fang, PhD; Jane E Alty MD, MA, MBChir.

Stefan Williams, University of Leeds, Leeds Institute of Health Science, Leeds, UK

Zhibin Zhao, (1) University of Manchester, Division of Informatics, Imaging & Data Sciences, Manchester, UK; (2) Xi’an Jiaotong University, School of Mechanical Engineering, Xi’an, China.

Awais Hafeez, (1) University of Leeds, School of Mechanical Engineering, Leeds, UK; (2) University of Engineering and Technology Lahore, Department of Mechatronics and Control Engineering, Lahore, Pakistan

David C Wong, University of Manchester, Division of Informatics, Imaging & Data Sciences, Manchester, UK

Samuel D Relton, University of Leeds, Leeds Institute of Health Science, Leeds, UK

Hui Fang, Loughborough University, Department of Computer Science, Loughborough, UK

Jane E Alty, (1) University of Tasmania, Wicking Dementia Research & Education Centre, Hobart, Australia (2) Leeds Teaching Hospitals NHS Trust UK.

Word count: 2758
Number of Tables: 3
Number of Figures: 3
Number of Videos: 1

Corresponding Author:
Stefan Williams
Leeds Institute of Health Science, 10th Floor, Worsley Building
University of Leeds, Clarendon Way, Leeds, UK LS2 9LU
Phone: 0113 343 0839
umswi@leeds.ac.uk
ORCID: 0000-0002-4786-1356

Keywords:
Parkinson’s disease; Parkinsonism; Bradykinesia; Finger tapping; Computer Vision; Artificial Intelligence
ABSTRACT

Objective
The worldwide prevalence of Parkinson’s disease is increasing. There is urgent need for new tools to objectively measure the condition. Existing methods to record the cardinal motor feature of the condition, bradykinesia, using wearable sensors or smartphone apps have not reached large-scale, routine use. We evaluate new computer vision (artificial intelligence) technology, DeepLabCut, as a contactless method to quantify measures related to Parkinson’s bradykinesia from smartphone videos of finger tapping.

Methods
Standard smartphone video recordings of 133 hands performing finger tapping (39 idiopathic Parkinson’s patients and 30 controls) were tracked on a frame-by-frame basis with DeepLabCut. Objective computer measures of tapping speed, amplitude and rhythm were correlated with clinical ratings made by 22 movement disorder neurologists using the Modified Bradykinesia Rating Scale (MBRS) and Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS).

Results
DeepLabCut reliably tracked and measured finger tapping in standard smartphone video. Computer measures correlated well with clinical ratings of bradykinesia (Spearman coefficients): -0.74 speed, 0.66 amplitude, -0.65 rhythm for MBRS; -0.56 speed, 0.61 amplitude, -0.50 rhythm for MDS-UPDRS; 0.69 combined for MDS-UPDRS. All p < 0.001.

Conclusion
New computer vision software, DeepLabCut, can quantify three measures related to Parkinson’s bradykinesia from smartphone videos of finger tapping. Objective ‘contactless’ measures of standard clinical examinations were not previously possible with wearable sensors (accelerometers, gyroscopes, infrared markers). DeepLabCut requires only conventional video recording of clinical examination and is entirely ‘contactless’. This next generation technology holds potential for Parkinson’s and other neurological disorders with altered movements.
INTRODUCTION

The worldwide prevalence of Parkinson’s is increasing, with an estimated 10 million people already affected [1]. There is urgent need to find new tools to objectively measure the condition, to assist diagnosis, monitoring, and research metrics [2]. This is especially pressing in the era of precision medicine and with a global shortage of neurologists [3].

The cornerstone of Parkinson’s assessment is the clinician’s visual judgement of bradykinesia: one of the cardinal motor features of the condition, which is defined as slowness of movement and decrement in size or speed, or progressive hesitations/halts, as movements are continued [4]. There are several clinical examination methods to assess bradykinesia, but in a brief clinic room assessment, one of the most common is to observe the patient repeatedly tapping their index finger against thumb ‘as quickly and as big as possible’ [5]. However, clinicians’ visual interpretation is inherently subjective [6,7], with considerable inter-rater variability [5,9-11]. Previous reported methods to measure bradykinesia on finger tapping have required wearable equipment such as gyroscopes [9], electromagnetic sensors [12; 13], and infrared camera markers [14], or patient interaction with an app-based task [15], and this has hindered the pathway to large-scale, routine clinical use.

Recent developments in computer interpretation of video (computer vision), clear the pathway to ‘contactless’ methods of quantifying clinical assessments [16], with potential widespread global application because of the availability of inexpensive smartphone cameras [17]. DeepLabCut is recently published open source computer vision software, without a requirement to write new computer code, that has been demonstrated to track and measure mice and fruit fly body parts [18,19]. We apply it to clinical neurology for the first time and assess whether the technology can be used to track and measure bradykinesia in standard smartphone video of conventional finger tapping examination.
METHODS

Ethical review

The study was approved by the North of Scotland Research Ethics Committee, United Kingdom Health Research Authority (IRAS project ID 256116).

Participants and video recording

Video recordings of 138 hands (including left and right) were obtained from 39 patients with idiopathic Parkinson’s disease and 30 healthy controls, who gave written consent. One video was rejected because the hand moved outside the video frame, making 137 videos: 77 Parkinson’s hands and 60 control hands.

All Parkinson’s patients had previously been diagnosed by a movement disorder specialist neurologist at Leeds Teaching Hospitals NHS Trust United Kingdom, according to Movement Disorder Society clinical diagnostic criteria [4], and were subjectively and objectively in the ‘on’ state at the time of video recording (no medication was withheld prior to recording). We did not exclude patients with postural hand tremor or dystonia. Controls were recruited from the companions of patients, or hospital/university staff, and had no history of Parkinson’s disease or other neurological diagnosis.

Participants rested their elbow on a chair arm with the forearm lifted to a 45 degree angle. The hand was free to move as per the protocol of the Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) Item 3.4 Finger Tapping [5]. Only the hand/forearm was within the video frame. The distance from camera to hand was approximately 1m, but not tightly defined. Digits 1 and 2 were closest to the camera. No specific instructions were given
for the position of digits 3 to 5, and participants were free to position these digits as they preferred, although the researcher gave a brief demonstration in which digits 3 to 5 were extended.

A smartphone, placed on a tripod, was used to record standard video (60 frames per second, 1920x1080 pixel resolution), with only ambient lighting. Participants were instructed to tap their index finger and thumb together “as quickly and as big as possible” for at least 10 seconds [5]. Each video was edited to an 11 second clip: 1 second prior to tapping onset and 10 seconds of finger tapping.

Clinical rating

22 consultant neurologists specialising in movement disorders (United Kingdom) were asked to rate 30 tapping videos each, selected at random from the set of 137 videos. The raters were blinded to patient / healthy control status and each other’s scores. Each video was rated according to both the MDS-UPDRS Item 3.4 Finger Tapping [5], and the Modified Bradykinesia Rating Scale (MBRS) [8,9]. The MDS-UPDRS requires the rater to amalgamate judgments of finger tapping speed, amplitude, and rhythm into a single composite score (Table 1). In contrast, the MBRS is comprised of three separate scores for speed, amplitude, and rhythm (Table 2). The modal clinician rating for each video was used for correlation with computer measures (see below). For each possible pair of raters, Cohen Kappa was calculated based on all videos they had both seen. The resulting values had mean 0.28 (fair agreement) and standard deviation 0.28.

DeepLabCut video tracking

DeepLabCut tracks the geometrical configuration of multiple body parts in video, without a requirement to wear markers [18,20]. It uses transfer learning, with feature detectors based on deep neural networks that have been pretrained on ImageNet, a massive object recognition dataset [21].
This means that DeepLabCut is able to accurately recognise and track body parts with minimal training data [18].

The set of 137 finger tapping videos was processed by DeepLabCut. We localised (labelled) six distinct points on the tapping hand in 20 frames selected by k-means clustering from each 660 frame video: thumb tip; index finger tip; thumb metacarpophalangeal (MCP) joint; index finger MCP joint; middle finger tip; dorsal wrist/proximal dorsal hand. The deep neural network architecture of DeepLabCut was then trained using these points to predict their localisation in the remaining 97% of (unlabelled) video frames (1030000 training iterations, ResNet 50). This created video pixel coordinates for finger tip and thumb tip throughout each video.

The accuracy of DeepLabCut to track the hand localisation points (including finger tip and thumb tip) was assessed using the ‘evaluate network’ function within DeepLabCut. This function computes the Euclidean error between the manual labels and the ones predicted by DeepLabCut averaged over the hand locations and test images (mean absolute error, proportional to the average root mean square error) [20].

Signal processing

Video pixel coordinates for the labels produced by DeepLabCut were used to calculate the pixel distance between index finger tip and thumb tip. A Savitzky-Golay filter was applied to the resulting time series, which removed large, sudden transient label ‘jumps’ caused by DeepLabCut mislabelling (i.e. large-distance, physiologically impossible label movement across one pair of video frames). The pixel number distances were standardised across all videos, by using the maximum opening distance between finger and thumb tip and normalising this to a value of 1 (all values were divided by the maximal value in the corresponding time series).
Three features of the resulting finger tip to thumb tip distance time series were calculated to reflect the clinical features of finger tapping speed, amplitude and rhythm (Figure 1). A measure of speed was calculated as the mean rate of change of the normalised distance between finger and thumb tip over time. A measure of amplitude variability was calculated by dividing each time series into one-second windows, with maximal overlap, and then calculating the coefficient of variation of the mean difference between the maximum and minimum amplitudes in each window. A measure of rhythm regularity was calculated by undertaking Fast Fourier Transform to find the distribution of frequencies within each finger tap time series, and then measuring the power of the dominant frequency peak added to the power of the frequencies 0.2 Hz either side of it. A more regular tapping rhythm concentrates power in a narrow frequency band, increasing the power of the dominant frequency peak (and its immediate neighbours), whereas a more irregular tapping rhythm leads to a more widely spread distribution of frequency bands, reducing the power of the dominant frequency peak (and its immediate neighbours).

Spearman correlation coefficients were calculated for the each of the three DeepLabCut measures vs the modal MBRS and MDS-UPDRS clinician ratings. In addition, we calculated the Spearman correlation coefficient of the three computer scores combined (the normalised arithmetic mean of the computer speed score, amplitude variability score and rhythm regularity score) with the MDS-UPDRS rating.
RESULTS

The participant and hand video characteristics are given in Table 3.

Tracking accuracy

DeepLabCut reliably tracked the finger tapping movements across 133 videoed clinical examinations. Figure 2 shows example frames from a video labelled by DeepLabCut. The mean absolute error of DeepLabCut labelling was 8.39 pixels, i.e. the average distance between manual (human) labels and those predicted by DeepLabCut was 8.39 pixels within a 1920x1080 pixel video frame. Video 1 shows examples of four videos labelled by DeepLabCut, including particularly challenging cases with tremor and dystonia also present.

Correlation with clinical bradykinesia ratings

The quantitative scores derived from DeepLabCut measurements correlated well with the group neurologists’ modal ratings of bradykinesia; Figure 3. The computer vision measures of tapping speed, amplitude variability and rhythm regularity had good correlations with the respective MBRS clinical ratings for speed (-0.70, p < 0.001), amplitude (0.65, p < 0.001) and rhythm (-0.61, p < 0.001). There were also good correlations between the computer measures of tapping speed (-0.66, p < 0.001),
amplitude variability (0.56, p < 0.001) and rhythm regularity (-0.50, p < 0.001) and the MDS-UPDRS clinical rating. The MDS-UPDRS is a composite clinical rating, combining the separate components of bradykinesia, and the three computer measures combined also correlated with MDS-UPDRS (-0.69, p < 0.001).

**DISCUSSION**

We have shown that new computer vision software, DeepLabCut, can track finger tapping in a standard smartphone video recording of a conventional clinical examination, and quantify three measures related to Parkinson’s bradykinesia: speed, amplitude variation and rhythm regularity. We found good correlations between the objective computer measures of finger tapping and clinical ratings of bradykinesia by 22 movement disorder neurologists using two gold standard clinical scales (MBRS and MDS-UPDRS). Thus, smartphone video measures using DeepLabCut exhibit convergent validity with conventional clinician ratings.

These results are new and clinically significant. This is the first study to apply new open source DeepLabCut technology, developed for animal behaviour studies, to the field of human movement disorders, or indeed any aspect of clinical neurology. We demonstrate new objective measures related to bradykinesia by bringing together two ubiquitous ingredients: the smartphone video camera and visual artificial intelligence software. With the worldwide ownership of smartphones so high (e.g. 82% in the UK, 77% in the US) [17], there are no hardware barriers to further use of this approach.

Neurologists recognize that quantification of bradykinesia will improve Parkinson’s management through ‘precision medicine’ principles: it has the potential to aid early accurate diagnosis and remote monitoring, and provide fine grain regular measurement of clinical trial outcomes. Previous methods
to quantify bradykinesia have focussed on wearable sensors measuring finger tapping [9,12-14,23] but it has been challenging to incorporate these into widespread use because equipment needs to be physically attached to the patient. Furthermore, the necessity for sensors to stay fixed on a specific body part has often required the exclusion of tremulous and dyskinetic patients [8,9,12,14,23]. In contrast, the DeepLabCut test presented here is a contactless method that simply requires a short 10 second video recording of the standard clinical examination. This means it can potentially be used in any setting and any patient.

The ‘real life’ versatility of DeepLabCut in ordinary clinical settings was evaluated in this study because no special efforts were made to optimize participant positioning or filming conditions. The tracking and measures were robust despite variations in smartphone camera distance, positioning of the hand, skin color, flexion / extension postures of digits 3 to 5, ambient lighting conditions, and inclusion of a small number of patients with tremor or dystonia during tapping. Video 1 illustrates this, with accurate labelling of finger and thumb tip despite superimposed tremor or dystonia. This is particularly encouraging for the relevance of DeepLabCut aiding assessment of Parkinson’s in conventional clinical settings.

Two further strengths of our study are the collection of clinical ratings from a large group of movement disorder neurologists, and the use of two rating scales: the MBRS and MDS-UPDRS, such that the clinical ratings were representative and reliable. Clinical rating scales are vulnerable to inter-rater variability but using modal values derived from a large number of blinded clinical raters, and two different gold-standard rating scales, provided a more robust ‘ground truth’ of clinical bradykinesia to evaluate the new test against. As such we were able to demonstrate that clinician and DeepLabCut measures correlate for both individual components of bradykinesia (MBRS) and an overall combination of those components (MDS-UPRDS).

Only three previous studies have objectively measured finger tapping using contactless, standard video analysis. One method extracted finger tapping information from 13 Parkinson's patients with
advanced disease [24], but it required inclusion of the patient's face in the video, limiting the practical utility. The second report, from our own group, described a video method to measure movement using an optical flow field [25] that interprets movements of the whole hand. The work here improves on this by explicitly extracting the salient thumb and finger tip points, and validating in a larger cohort. A third study of 60 Parkinson’s patients (no controls) found strong correlation between MDS-UPRDS finger tap rating and video measures of tap interval (frequency) ($r=0.91$), frequency variation ($r=0.82$) and amplitude ($r=0.94$) but not amplitude variation ($r=0.39$) [26]. The ‘ground truth’ of the clinical comparison was less robust, based on just two clinical raters, and there were no MBRS ratings; furthermore, some images were blurred in the 24 frame per second videos. A major advantage of our method in comparison is that DeepLabCut is open source software, that can be downloaded and used without the need to write computer code, making it available now to neurologists with only limited computing knowledge.

The present study has some limitations. Like human vision, a simple camera lacks an absolute measure of distance, and our method cannot capture 3D movement. Rotation of the thumb and finger into a horizontal plane would falsely alter amplitude measurement. However, in practice such movement is rare during tapping, and our normalized measure of finger-thumb distance based on 2D maximum opening distance appeared to capture amplitude, speed and rhythm abnormalities well, given the good correlations with clinical ratings.

We have measured amplitude variation, but not amplitude decrement (the sequence effect). Previous approaches have measured decrement as a straight line linear regression [23], but decrement would also proportionately affect a measure of amplitude variation, such that the variations in decrement should be captured within our measure. Furthermore, there are several other previous reports of good correlation between non-decrement amplitude measures using sensors and clinician ratings of finger tap bradykinesia [9].
Relatively few videos were rated as grade 3 or 4 bradykinesia, but this would not be expected to greatly affect our findings, as computer analysis of milder bradykinesia could be considered a more challenging task. There was a degree of error in DeepLabCut labelling of finger and thumb tips, but this was minimal (mean of 8.39 pixels within a 1920x1080 pixel video frame) and unlikely to affect individual measures. We mitigated against this by applying a filter prior to feature extraction. Full flexion of digits 3 to 5 might improve labelling accuracy, but would also limit the amplitude of finger tapping [23], potentially masking more subtle deficits. The labelling error provided by DeepLabCut’s ‘evaluate network’ function only gives an accuracy figure for the entire set of videos but not for individual videos, meaning that we have not reported accuracy in subsets of videos, such as those with visible tremor.

Finally, we have demonstrated measurements that show convergent validity with clinical ratings, rather than a test to discriminate patients from controls. In our view, it makes little sense to pursue automated diagnostic classification using finger tapping video alone because clinical Parkinson’s diagnosis is by definition a broader assessment than isolated finger tapping performance [4].

This study has evaluated the clinical application of DeepLabCut in a medical condition, specifically a neurological disorder. The methodology described here provides a new objective measure of Parkinson’s bradykinesia in conventional clinical examination. DeepLabCut provides a contactless means for tracking and measuring bradykinesia in Parkinson’s, without any special equipment other than the ubiquitous smartphone. This research therefore provides a new tool to quantify bradykinesia. It could potentially be used to support diagnosis and monitoring of Parkinson’s, both clinically and as an outcome measure in research studies. Beyond Parkinson’s and other movement disorders, this same technology could be applied to many other neurological conditions that require measurement of movements or function; for example multiple sclerosis, neuropathies, epilepsy and stroke.
REFERENCES


Conflict of Interest Statement

The authors declare that they have no conflict of interest.

FIGURE LEGENDS

Figure 1. Illustration of the three parameters derived from the DeepLabCut finger tip and thumb tip coordinates to measure finger tapping speed, amplitude and rhythm.

Figure 2. Example video frames taken from smartphone video labelled by DeepLabCut, showing the six localization (labelling) points: thumb tip (dark blue), thumb metacarpophalangeal (MCP) joint (light blue), index finger tip (cyan), index finger MCP joint (yellow), middle finger tip (orange), dorsal wrist (red).

Figure 3. Computer measures of finger tapping speed, amplitude variation and rhythm, derived from DeepLabCut smartphone video tracking, correlate with clinician mode ratings for the same videos. Pearson correlation coefficients shown. MBRS: Modified Bradykinesia Rating Scale. MDS-UPDRS: Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale.
Table 1. The Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) Item 3.4 (Finger Tapping).

<table>
<thead>
<tr>
<th>MDS-UPDRS Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: Normal</td>
<td>No problems</td>
</tr>
<tr>
<td>1: Slight</td>
<td>Any of the following: a) the regular rhythm is broken with one or two interruptions or hesitations of the tapping movement; b) slight slowing; c) the amplitude decrements near the end of the 10 taps.</td>
</tr>
<tr>
<td>2: Mild</td>
<td>Any of the following: a) 3 to 5 interruptions during tapping; b) mild slowing; c) the amplitude decrements midway in the 10-tap sequence.</td>
</tr>
<tr>
<td>3: Moderate</td>
<td>Any of the following: a) more than 5 interruptions during tapping or at least one longer arrest (freeze) in ongoing movement; b) moderate slowing; c) the amplitude decrements starting after the 1st tap.</td>
</tr>
<tr>
<td>4: Severe</td>
<td>Cannot or can only barely perform the task because of slowing, interruptions, or decrements.</td>
</tr>
</tbody>
</table>

Each hand is tested separately. The patient is instructed to tap the index finger on the thumb “as quickly and as big as possible”. The first 10 taps are evaluated by the clinician [5].

Table 2. The Modified Bradykinesia Rating Scale (MBRS).
Table 3. Study participant (hand video) characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Patients</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Std. Dev.) yrs</td>
<td>68 (9.6)</td>
<td>59 (19.4)</td>
</tr>
<tr>
<td>Male/Female</td>
<td>47/26</td>
<td>22/38</td>
</tr>
<tr>
<td>Median years since diagnosis</td>
<td>4</td>
<td>n/a</td>
</tr>
<tr>
<td>Median H&amp;Y [IQR]</td>
<td>2 [1,3]</td>
<td>n/a</td>
</tr>
<tr>
<td>H&amp;Y = 1</td>
<td>32</td>
<td>2</td>
</tr>
<tr>
<td>H&amp;Y = 1.5</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>H&amp;Y = 2</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>H&amp;Y = 2.5</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>H&amp;Y = 3</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>H&amp;Y = 4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>H&amp;Y = 5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Median MDS-UPDRS [IQR]</td>
<td>2 [1,3]</td>
<td>1 [0,1]</td>
</tr>
<tr>
<td>MDS-UPDRS = 0</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td>MDS-UPDRS = 1</td>
<td>20</td>
<td>23</td>
</tr>
</tbody>
</table>
MDS-UPDRS = 2 19 12
MDS-UPDRS = 3 20 2
MDS-UPDRS = 4 5 0
Visible tremor in video 11 0
Visible dystonia in video 2 0

Data is split by Parkinson’s hands (n=73) and control hands (n=60). H&Y: modified Hoehn and Yahr scale [22]. MDS-UPDRS: Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale, Item 3.4 (Finger Tapping). IQR: Interquartile Range.

Acknowledgements

We would like to thank the following UK movement disorder neurologists for providing clinical rating of videos: Dr Jeremy Cosgrove, Dr Rachel Newby, Dr Stephen Butterworth, Dr Alistair Lansbury, Dr Michael Bonello, Dr Sundus Alusi, Dr Peter Moore, Dr Richard Ellis, Dr Camille Carroll, Dr Chris Kobylecki, Dr Paul Goldsmith, Dr Naomi Warren, Dr Edward Newman, Dr Philip Buttery, Dr Jonathan Evans, Dr Gillian Sare, Dr Donald Grosset, Dr Caroline Williams-Gray, Dr Monty Silverdale, Dr Louise Wiblin, Dr Adam Cassidy

We would also like to thank Dr Parisa Patel and Dr Paschal O’Gorman for data collection.

Author contributions

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Role</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stefan Williams, MBBS, BSc</td>
<td>University of Leeds, UK</td>
<td>Author</td>
<td>Design and conceptualized study; data acquisition and interpretation;</td>
</tr>
<tr>
<td>Name</td>
<td>Contributions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhibin Zhao, BEng</td>
<td>Analyzed the data (major role)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) University of Manchester, UK; (2) Xi’an Jiaotong University, China</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awais Hafeez, MA, BSc</td>
<td>Design and conceptualized the study; major role in acquisition of data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) University of Leeds, UK; (2) University of Engineering and Technology Lahore, Pakistan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>David C Wong, PhD</td>
<td>Design and conceptualized study; data interpretation; revised the manuscript for intellectual content</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>University of Manchester, UK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Samuel D Relton, PhD</td>
<td>Design and conceptualized study; data interpretation; revised the manuscript for intellectual content</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>University of Leeds, UK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hui Fang, PhD</td>
<td>Design and conceptualized study (major role)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Loughborough University, UK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jane E Alty, MD, MA, MBBChir</td>
<td>Design and conceptualized study; data interpretation; drafted the manuscript for intellectual content</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) University of Tasmania, Australia; (2) Leeds Teaching Hospitals NHS Trust, UK</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The discerning eye of computer vision: can it measure Parkinson’s finger tap bradykinesia?

Authors: Stefan Williams, MBBS, BSc; Zhibin Zhao, BEng; Awais Hafeez, MS, BSc; David C Wong, PhD; Samuel D Relton, PhD; Hui Fang, PhD; Jane E Alty MD, MA, MBBChir.

Stefan Williams, University of Leeds, Leeds Institute of Health Science, Leeds, UK

Zhibin Zhao, (1) University of Manchester, Division of Informatics, Imaging & Data Sciences, Manchester, UK; (2) Xi’an Jiatong University, School of Mechanical Engineering, Xi’an, China.

Awais Hafeez, (1) University of Leeds, School of Mechanical Engineering, Leeds, UK; (2) University of Engineering and Technology Lahore, Department of Mechatronics and Control Engineering, Lahore, Pakistan

David C Wong, University of Manchester, Division of Informatics, Imaging & Data Sciences, Manchester, UK

Samuel D Relton, University of Leeds, Leeds Institute of Health Science, Leeds, UK

Hui Fang, Loughborough University, Department of Computer Science, Loughborough, UK

Jane E Alty, (1) University of Tasmania, Wicking Dementia Research & Education Centre, Hobart, Australia (2) Leeds Teaching Hospitals NHS Trust UK.

Word count: 3101
Number of Tables: 3
Number of Figures: 3
Number of Videos: 1

Corresponding Author:
Stefan Williams
Leeds Institute of Health Science, 10th Floor, Worsley Building
University of Leeds, Clarendon Way, Leeds, UK LS2 9LU
Phone: 0113 343 0839
umswi@leeds.ac.uk
ORCID: 0000-0002-4786-1356

Keywords:
Parkinson’s disease; Parkinsonism; Bradykinesia; Finger tapping; Computer Vision; Artificial Intelligence; DeepLabCut
ABSTRACT

Objective
The worldwide prevalence of Parkinson’s disease is increasing. There is urgent need for new tools to objectively measure the condition. Existing methods to record the cardinal motor feature of the condition, bradykinesia, using wearable sensors or smartphone apps have not reached large-scale, routine use. We evaluate new computer vision (artificial intelligence) technology, DeepLabCut, as a contactless method to quantify measures related to Parkinson’s bradykinesia from smartphone videos of finger tapping.

Methods
Standard smartphone video recordings of 133 hands performing finger tapping (39 idiopathic Parkinson’s patients and 30 controls) were tracked on a frame-by-frame basis with DeepLabCut. Objective computer measures of tapping speed, amplitude and rhythm were correlated with clinical ratings made by 22 movement disorder neurologists using the Modified Bradykinesia Rating Scale (MBRS) and Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS).

Results
DeepLabCut reliably tracked and measured finger tapping in standard smartphone video. Computer measures correlated well with clinical ratings of bradykinesia (Spearman coefficients): -0.74 speed, 0.66 amplitude, -0.65 rhythm for MBRS; -0.56 speed, 0.61 amplitude, -0.50 rhythm for MDS-UPDRS; 0.69 combined for MDS-UPDRS. All p < 0.001.

Conclusion
New computer vision software, DeepLabCut, can quantify three measures related to Parkinson’s bradykinesia from smartphone videos of finger tapping. Objective ‘contactless’ measures of standard clinical examinations were not previously possible with wearable sensors (accelerometers, gyroscopes, infrared markers). DeepLabCut requires only conventional video recording of clinical examination and is entirely ‘contactless’. This next generation technology holds potential for Parkinson’s and other neurological disorders with altered movements.
INTRODUCTION

The worldwide prevalence of Parkinson’s is increasing, with an estimated 10 million people already affected [1]. There is urgent need to find new tools to objectively measure the condition, to assist diagnosis, monitoring, and research metrics [2]. This is especially pressing in the era of precision medicine and with a global shortage of neurologists [3].

The cornerstone of Parkinson’s assessment is the clinician’s visual judgement of bradykinesia: one of the cardinal motor features of the condition, which is defined as slowness of movement and decrement in size or speed, or progressive hesitations/halts, as movements are continued [4]. There are several clinical examination methods to assess bradykinesia, but in a brief clinic room assessment, one of the most common is to observe the patient repeatedly tapping their index finger against thumb ‘as quickly and as big as possible’ [5]. However, clinicians’ visual interpretation is inherently subjective [6,7], with considerable inter-rater variability [5,9-11]. Previous reported methods to measure bradykinesia on finger tapping have required wearable equipment such as gyroscopes [9], electromagnetic sensors [12; 13], and infrared camera markers [14], or patient interaction with an app-based task [15], and this has hindered the pathway to large-scale, routine clinical use.

Recent developments in computer interpretation of video (computer vision), clear the pathway to ‘contactless’ methods of quantifying clinical assessments [16], with potential widespread global application because of the availability of inexpensive smartphone cameras [17]. DeepLabCut is recently published open source computer vision software, without a requirement to write new computer code, that has been demonstrated to track and measure mice and fruit fly body parts [18,19]. We apply it to clinical neurology for the first time and assess whether the technology can be used to track and measure bradykinesia in standard smartphone video of conventional finger tapping examination.

METHODS
Ethical review

The study was approved by the North of Scotland Research Ethics Committee, United Kingdom Health Research Authority (IRAS project ID 256116).

Participants and video recording

Video recordings of 138 hands (including left and right) were obtained from 39 patients with idiopathic Parkinson’s disease and 30 healthy controls, who gave written consent. One video was rejected because the hand moved outside the video frame, making 137 videos: 77 Parkinson’s hands and 60 control hands.

All Parkinson’s patients had previously been diagnosed by a movement disorder specialist neurologist at Leeds Teaching Hospitals NHS Trust United Kingdom, according to Movement Disorder Society clinical diagnostic criteria [4], and were subjectively and objectively in the ‘on’ state at the time of video recording (no medication was withheld prior to recording). We did not exclude patients with postural hand tremor or dystonia. Controls were recruited from the companions of patients, or hospital/university staff, and had no history of Parkinson’s disease or other neurological diagnosis.

Participants rested their elbow on a chair arm with the forearm lifted to a 45 degree angle. The hand was free to move as per the protocol of the Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) Item 3.4 Finger Tapping [5]. Only the hand/forearm was within the video frame. The distance from camera to hand was approximately 1m, but not tightly defined. Digits 1 and 2 were closest to the camera. No specific instructions were given for the position of digits 3 to 5, and participants were free to position these digits as they preferred, although the researcher gave a brief demonstration in which digits 3 to 5 were extended.
A smartphone, placed on a tripod, was used to record standard video (60 frames per second, 1920x1080 pixel resolution), with only ambient lighting. Participants were instructed to tap their index finger and thumb together “as quickly and as big as possible” for at least 10 seconds [5]. Each video was edited to an 11 second clip: 1 second prior to tapping onset and 10 seconds of finger tapping.

Clinical rating

22 consultant neurologists specialising in movement disorders (United Kingdom) were asked to rate 30 tapping videos each, selected at random from the set of 137 videos. The median number of raters per video was 5 (range 1 to 12, interquartile range 3 to 7). The raters were blinded to patient / healthy control status and each other’s scores. Each video was rated according to both the MDS-UPDRS Item 3.4 Finger Tapping [5], and the Modified Bradykinesia Rating Scale (MBRS) [8,9]. The MDS-UPDRS requires the rater to amalgamate judgments of finger tapping speed, amplitude, and rhythm into a single composite score (Table 1). In contrast, the MBRS is comprised of three separate scores for speed, amplitude, and rhythm (Table 2). The modal clinician rating for each video was used for correlation with computer measures (see below). For each possible pair of raters, Cohen Kappa was calculated based on all videos they had both seen. The resulting values had mean 0.28 (fair agreement) and standard deviation 0.28.

DeepLabCut video tracking

DeepLabCut tracks the geometrical configuration of multiple body parts in video, without a requirement to wear markers [18,20]. It uses transfer learning, with feature detectors based on deep neural networks that have been pretrained on ImageNet, a massive object recognition dataset [21]. This means that DeepLabCut is able to accurately recognise and track body parts with minimal training data [18].

The set of 137 finger tapping videos was processed by DeepLabCut. We localised (labelled) six distinct points on the tapping hand in 20 frames selected by k-means clustering from each 660 frame video:
thumb tip; index finger tip; thumb metacarpophalangeal (MCP) joint; index finger MCP joint; middle finger tip; dorsal wrist/proximal dorsal hand. The deep neural network architecture of DeepLabCut was then trained using these points to predict their localisation in the remaining 97% of (unlabelled) video frames (1030000 training iterations, ResNet 50). This created video pixel coordinates for finger tip and thumb tip throughout each video.

The accuracy of DeepLabCut to track the hand localisation points (including finger tip and thumb tip) was assessed using the ‘evaluate network’ function within DeepLabCut. This function computes the Euclidean error between the manual labels and the ones predicted by DeepLabCut averaged over the hand locations and test images (mean absolute error, proportional to the average root mean square error) [20].

Signal processing

Video pixel coordinates for the labels produced by DeepLabCut were used to calculate the pixel distance between index finger tip and thumb tip. A Savitzky-Golay filter was applied to the resulting time series, which removed large, sudden transient label ‘jumps’ caused by DeepLabCut mislabelling (i.e. large-distance, physiologically impossible label movement across one pair of video frames). The pixel number distances were standardised across all videos, by using the maximum opening distance between finger and thumb tip and normalising this to a value of 1 (all values were divided by the maximal value in the corresponding time series).

Three features of the resulting finger tip to thumb tip distance time series were calculated to reflect the clinical features of finger tapping speed, amplitude and rhythm (Figure 1). A measure of speed was calculated as the mean rate of change of the normalised distance between finger and thumb tip over time. A measure of amplitude variability was calculated by dividing each time series into one-second windows, with maximal overlap, and then calculating the coefficient of variation of the mean difference between the maximum and minimum amplitudes in each window. A measure of rhythm regularity was
calculated by undertaking Fast Fourier Transform to find the distribution of frequencies within each finger tap time series, and then measuring the power of the dominant frequency peak added to the power of the frequencies 0.2 Hz either side of it. A more regular tapping rhythm concentrates power in a narrow frequency band, increasing the power of the dominant frequency peak (and its immediate neighbours), whereas a more irregular tapping rhythm leads to a more widely spread distribution of frequency bands, reducing the power of the dominant frequency peak (and its immediate neighbours).

Spearman correlation coefficients were calculated for the each of the three DeepLabCut measures vs the modal MBRS and MDS-UPDRS clinician ratings. In addition, we calculated the Spearman correlation coefficient of the three computer scores combined (the normalised arithmetic mean of the computer speed score, amplitude variability score and rhythm regularity score) with the MDS-UPDRS rating.

RESULTS

The participant and hand video characteristics are given in Table 3.

Tracking accuracy

DeepLabCut reliably tracked the finger tapping movements across 133 videoed clinical examinations. Figure 2 shows example frames from a video labelled by DeepLabCut. The mean absolute error of DeepLabCut labelling was 8.39 pixels, i.e. the average distance between manual (human) labels and those predicted by DeepLabCut was 8.39 pixels within a 1920x1080 pixel video frame. Video 1 shows examples of four videos labelled by DeepLabCut, including particularly challenging cases with tremor and dystonia also present.

Correlation with clinical bradykinesia ratings
The quantitative scores derived from DeepLabCut measurements correlated well with the group neurologists’ modal ratings of bradykinesia; Figure 3. The computer vision measures of tapping speed, amplitude variability and rhythm regularity had good correlations with the respective MBRS clinical ratings for speed (-0.70, p < 0.001), amplitude (0.65, p < 0.001) and rhythm (-0.61, p < 0.001). There were also good correlations between the computer measures of tapping speed (-0.66, p < 0.001), amplitude variability (0.56, p < 0.001) and rhythm regularity (-0.50, p < 0.001) and the MDS-UPDRS clinical rating. The MDS-UPDRS is a composite clinical rating, combining the separate components of bradykinesia, and the three computer measures combined also correlated with MDS-UPDRS (-0.69, p < 0.001).

DISCUSSION

We have shown that new computer vision software, DeepLabCut, can track finger tapping in a standard smartphone video recording of a conventional clinical examination, and quantify three measures related to Parkinson’s bradykinesia: speed, amplitude variation and rhythm regularity. We found good correlations between the objective computer measures of finger tapping and clinical ratings of bradykinesia by 22 movement disorder neurologists using two gold standard clinical scales (MBRS and MDS-UPDRS). Thus, smartphone video measures using DeepLabCut exhibit convergent validity with conventional clinician ratings.

These results are new and clinically significant. This is the first study to apply new open source DeepLabCut technology, developed for animal behaviour studies, to the field of human movement disorders, or indeed any aspect of clinical neurology. We demonstrate new objective measures related to bradykinesia by bringing together two ubiquitous ingredients: the smartphone video camera and visual artificial intelligence software. With the worldwide ownership of smartphones so high (e.g. 82% in the UK, 77% in the US) [17], there are no hardware barriers to further use of this approach.
Neurologists recognize that quantification of bradykinesia will improve Parkinson’s management through ‘precision medicine’ principles: it has the potential to aid early accurate diagnosis and remote monitoring, and provide fine grain regular measurement of clinical trial outcomes. Previous methods to quantify bradykinesia have focussed on wearable sensors measuring finger tapping [9,12-14,23] but it has been challenging to incorporate these into widespread use because equipment needs to be physically attached to the patient. Furthermore, the necessity for sensors to stay fixed on a specific body part has often required the exclusion of tremulous and dyskinetic patients [8,9,12,14,23]. In contrast, the DeepLabCut test presented here is a contactless method that simply requires a short 10 second video recording of the standard clinical examination. This means it can potentially be used in any setting and any patient.

The ‘real life’ versatility of DeepLabCut in ordinary clinical settings was evaluated in this study because no special efforts were made to optimize participant positioning or filming conditions. The tracking and measures were robust despite variations in smartphone camera distance, positioning of the hand, skin color, flexion / extension postures of digits 3 to 5, ambient lighting conditions, and inclusion of a small number of patients with tremor or dystonia during tapping. Video 1 illustrates this, with accurate labelling of finger and thumb tip despite superimposed tremor or dystonia. This is particularly encouraging for the relevance of DeepLabCut aiding assessment of Parkinson’s in conventional clinical settings.

Two further strengths of our study are the collection of clinical ratings from a large group of movement disorder neurologists, and the use of two rating scales: the MBRS and MDS-UPDRS, such that the clinical ratings were representative and reliable. Clinical rating scales are vulnerable to inter-rater variability but using modal values derived from a large number of blinded clinical raters, and two different gold-standard rating scales, provided a more robust ‘ground truth’ of clinical bradykinesia to evaluate the new test against. As such we were able to demonstrate that clinician and DeepLabCut measures correlate for both individual components of bradykinesia (MBRS) and an overall combination.
of those components (MDS-UPRDS). An interesting future application with a larger participant group would be to examine the relative contribution of the three video measures to individual grades of UPDRS bradykinesia.

Another advantage of the approach described here is that it is not limited by the physical constraints of wearable or screen/keyboard tapping methods, and could be applied to detect and measure other motor features of Parkinson’s in a contactless manner such as tremor, gait, or posture. Computer vision is general has potential for application to these signs [24,25].

Only three previous studies have objectively measured finger tapping using contactless, standard video analysis. One method extracted finger tapping information from 13 Parkinson's patients with advanced disease [26], but it required inclusion of the patient's face in the video, limiting the practical utility. The second report, from our own group, described a video method to measure movement using an optical flow field [27] that interprets movements of the whole hand. The work here improves on this by explicitly extracting the salient thumb and finger tip points, and validating in a larger cohort. A third study of 60 Parkinson’s patients (no controls) found strong correlation between MDS-UPRDS finger tap rating and video measures of tap interval (frequency) (r=0.91), frequency variation (r=0.82) and amplitude (r=-0.94) but not amplitude variation (r=0.39) [28]. The ‘ground truth’ of the clinical comparison was less robust, based on just two clinical raters, and there were no MBRS ratings; furthermore, some images were blurred in the 24 frame per second videos. A major advantage of our method in comparison is that DeepLabCut is open source software, that can be downloaded and used without the need to write computer code, making it available now to neurologists with only limited computing knowledge.

A well-studied, non-camera method to measure finger tap bradykinesia is the ‘Bradykinesia-Akinesia Incoordination (BRAIN) Test’, which involves participants tapping a standard computer keyboard for 30 seconds [29]. This shares advantages with our method in that it involves no special equipment, and could be employed without clinician-patient contact. However, it requires patient motivation to engage
with a keyboard tapping task, rather than a simple recording of existing standard clinical examination (and similar requirements for patient motivation apply to patient smartphone apps [15]). Furthermore, ‘BRAIN’ correlations with MDS-UPDRS finger tapping ratings of 0.44, 0.20, 0.33, 0.03 [29], are weaker than those we report (0.66, 0.56, 0.50, 0.69). One reason for this may be that a computer keyboard cannot record any measure of tap amplitude, and the speed measure is limited to tap frequency, so that some core aspects of bradykinesia cannot be captured by ‘BRAIN’.

The present study has some limitations. Like human vision, a simple camera lacks an absolute measure of distance, and our method cannot capture 3D movement. Rotation of the thumb and finger into a horizontal plane would falsely alter amplitude measurement. However, in practice such movement is rare during tapping, and our normalized measure of finger-thumb distance based on 2D maximum opening distance appeared to capture amplitude, speed and rhythm abnormalities well, given the good correlations with clinical ratings.

We have measured amplitude variation, but not amplitude decrement (the sequence effect). Previous approaches have measured decrement as a straight line linear regression [23], but decrement would also proportionately affect a measure of amplitude variation, such that the variations in decrement should be captured within our measure. Furthermore, there are several other previous reports of good correlation between non-decrement amplitude measures using sensors and clinician ratings of finger tap bradykinesia [9].

Relatively few videos were rated as grade 3 or 4 bradykinesia, and it is possible that a greater number of higher grade videos might change the strength of correlations. Even with accurate tracking, all forms of rating or measuring bradykinesia (including clinician) may potentially be confounded by comorbidities such as joint deformities, pain, dystonia etc. However, with a larger dataset, machine learning techniques are well-suited to separate such additional contributions from underlying bradykinesia. There was a degree of error in DeepLabCut labelling of finger and thumb tips, but this was minimal (mean of 8.39 pixels within a 1920x1080 pixel video frame) and unlikely to affect
individual measures. We mitigated against this by applying a filter prior to feature extraction. Full flexion of digits 3 to 5 might improve labelling accuracy, but would also limit the amplitude of finger tapping [23], potentially masking more subtle deficits. The labelling error provided by DeepLabCut’s ‘evaluate network’ function only gives an accuracy figure for the entire set of videos but not for individual videos, meaning that we have not reported accuracy in subsets of videos, such as those with visible tremor. This study was exploratory, aimed at proof of concept that bradykinesia related measures could be derived from smartphone video. The protocol was thus relatively simple. The test-retest reliability of this technique has not been addressed in the study but we hope to do so in future. Similarly, we did not measure ‘off’ and ‘on’ medication states (sensitivity to change).

Finally, we have demonstrated measurements that show convergent validity with clinical ratings, rather than a test to discriminate patients from controls. In our view, it makes little sense to pursue automated diagnostic classification using finger tapping video alone because clinical Parkinson’s diagnosis is by definition a broader assessment than isolated finger tapping performance [4].

This study has evaluated the clinical application of DeepLabCut in a medical condition, specifically a neurological disorder. The methodology described here provides a new objective measure of Parkinson’s bradykinesia in conventional clinical examination. DeepLabCut provides a contactless means for tracking and measuring bradykinesia in Parkinson’s, without any special equipment other than the ubiquitous smartphone. This research therefore provides a new tool to quantify bradykinesia. It could potentially be used to support diagnosis and monitoring of Parkinson’s, both clinically and as an outcome measure in research studies. Beyond Parkinson’s and other movement disorders, this same technology could be applied to many other neurological conditions that require measurement of movements or function; for example multiple sclerosis, neuropathies, epilepsy and stroke.

REFERENCES


**Conflict of Interest Statement**

The authors declare that they have no conflict of interest.

**FIGURE LEGENDS**
Figure 1. Illustration of the three parameters derived from the DeepLabCut finger tip and thumb tip coordinates to measure finger tapping speed, amplitude and rhythm.

Figure 2. Example video frames taken from smartphone video labelled by DeepLabCut, showing the six localization (labelling) points: thumb tip (dark blue), thumb metacarpophalangeal (MCP) joint (light blue), index finger tip (cyan), index finger MCP joint (yellow), middle finger tip (orange), dorsal wrist (red).

Figure 3. Computer measures of finger tapping speed, amplitude variation and rhythm, derived from DeepLabCut smartphone video tracking, correlate with clinician mode ratings for the same videos. Pearson correlation coefficients shown. MBRS: Modified Bradykinesia Rating Scale. MDS-UPDRS: Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale.

TABLES

Table 1. The Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) Item 3.4 (Finger Tapping).

<table>
<thead>
<tr>
<th>MDS-UPDRS Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: Normal</td>
<td>No problems</td>
</tr>
<tr>
<td>1: Slight</td>
<td>Any of the following: a) the regular rhythm is broken with one or two interruptions or hesitations of the tapping movement; b) slight slowing; c) the amplitude decrements near the end of the 10 taps.</td>
</tr>
<tr>
<td>2: Mild</td>
<td>Any of the following: a) 3 to 5 interruptions during tapping; b) mild slowing; c) the amplitude decrements midway in the 10-tap sequence.</td>
</tr>
<tr>
<td>3: Moderate</td>
<td>Any of the following: a) more than 5 interruptions during tapping or at least one longer arrest (freeze) in ongoing movement; b) moderate slowing; c) the amplitude decrements starting after the 1st tap.</td>
</tr>
<tr>
<td>4: Severe</td>
<td>Cannot or can only barely perform the task because of slowing, interruptions, or decrements.</td>
</tr>
</tbody>
</table>
Each hand is tested separately. The patient is instructed to tap the index finger on the thumb "as quickly and as big as possible". The first 10 taps are evaluated by the clinician [5].

**Table 2.** The Modified Bradykinesia Rating Scale (MBRS).

<table>
<thead>
<tr>
<th>Score</th>
<th>Speed</th>
<th>Amplitude</th>
<th>Rhythm</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal</td>
<td>Normal</td>
<td>Regular, no arrests or pauses in ongoing movement</td>
</tr>
<tr>
<td>1</td>
<td>Mild Slowing</td>
<td>Mild reduction in amplitude in later performance, most movements close to normal</td>
<td>Mild impairment, up to two brief arrests in the 10 seconds, none lasting &gt; 1 second</td>
</tr>
<tr>
<td>2</td>
<td>Moderate Slowing</td>
<td>Moderate, reduction in amplitude visible early in performance but continues to maintain 50% amplitude through most of the tasks</td>
<td>Severe, 5 or more arrests/10 seconds; OR more than 2 lasting &gt; 1 second</td>
</tr>
<tr>
<td>3</td>
<td>Severe slowing</td>
<td>Severe, less than 50% amplitude through most of the task</td>
<td>Can barely perform the task</td>
</tr>
<tr>
<td>4</td>
<td>Can barely perform the task</td>
<td>Can barely perform the task</td>
<td>Can barely perform the task</td>
</tr>
</tbody>
</table>

Each hand is tested separately. The patient is instructed to tap the index finger on the thumb "as quickly and as big as possible". The first 10 seconds of tapping are evaluated by the clinician to give three MBRS ratings (one for each of speed, amplitude, rhythm) [6,7].

**Table 3.** Study participant (hand video) characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Patients</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Std. Dev.) yrs</td>
<td>68 (9.6)</td>
<td>59 (19.4)</td>
</tr>
<tr>
<td>Male/Female</td>
<td>47/26</td>
<td>22/38</td>
</tr>
<tr>
<td>Median years since diagnosis</td>
<td>4</td>
<td>n/a</td>
</tr>
<tr>
<td>Median H&amp;Y [IQR]</td>
<td>2 [1,3]</td>
<td>n/a</td>
</tr>
<tr>
<td>H&amp;Y = 1</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 1.5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 2</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 2.5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 3</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Median MDS-UPDRS [IQR]</td>
<td>2 [1,3]</td>
<td>1 [0,1]</td>
</tr>
<tr>
<td>MDS-UPDRS = 0</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td>MDS-UPDRS = 1</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>MDS-UPDRS = 2</td>
<td>19</td>
<td>12</td>
</tr>
<tr>
<td>MDS-UPDRS = 3</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>MDS-UPDRS = 4</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Visible tremor in video</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Visible dystonia in video</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Data is split by Parkinson's hands (n=73) and control hands (n=60). H&Y: modified Hoehn and Yahr scale [22]. MDS-UPDRS: Movement Disorder Society revision of the Unified Parkinson's Disease Rating Scale, Item 3.4 (Finger Tapping). IQR: Interquartile Range.

Acknowledgements

We would like to thank the following UK movement disorder neurologists for providing clinical rating of videos: Dr Jane Alty, Dr Sundus Alusi, Dr Michael Bonello, Dr Stephen Butterworth, Dr Philip Buttery, Dr Camille Carroll, Dr Adam Cassidy, Dr Jeremy Cosgrove, Dr Richard Ellis, Dr Jonathan Evans, Dr Paul Goldsmith, Dr Donald Grosset, Dr Christopher Kobylecki, Dr Alistair Lansbury, Dr Peter Moore, Dr Rachel Newby, Dr Edward Newman, Dr Gillian Sare, Dr Monty Silverdale, Dr Naomi Warren, Dr Louise Wiblin, Dr Caroline Williams-Gray.

We would also like to thank Dr Parisa Patel and Dr Paschal O’Gorman for data collection.

Author contributions

| Name | Location | Role | Contribution |
|------|----------|------|--------------|--------------|

C:\pdfbuilder\temp\Second_Revised(MARKED)ComputerVisionParkinsonsBradykinesia.docx
<table>
<thead>
<tr>
<th>Name</th>
<th>Institution/University</th>
<th>Authorship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stefan Williams, MBBS, BSc</td>
<td>University of Leeds, UK</td>
<td>Design and conceptualized study; data acquisition and interpretation; drafted the manuscript for intellectual content</td>
</tr>
<tr>
<td>Zhibin Zhao, BEng</td>
<td>(1) University of Manchester, UK; (2) Xi’an Jiaotong University, China</td>
<td>Analyzed the data (major role)</td>
</tr>
<tr>
<td>Awais Hafeez, MA, BSc</td>
<td>(1) University of Leeds, UK; (2) University of Engineering and Technology Lahore, Pakistan</td>
<td>Design and conceptualized the study; major role in acquisition of data</td>
</tr>
<tr>
<td>David C Wong, PhD</td>
<td>University of Manchester, UK</td>
<td>Design and conceptualized study; analyzed the data; data interpretation; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Samuel D Relton, PhD</td>
<td>University of Leeds, UK</td>
<td>Design and conceptualized study; data interpretation; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Hui Fang, PhD</td>
<td>Loughborough University, UK</td>
<td>Design and conceptualized study (major role)</td>
</tr>
<tr>
<td>Jane E Alty, MD, MA, MBBChir</td>
<td>(1) University of Tasmania, Australia; (2) Leeds Teaching Hospitals NHS Trust, UK</td>
<td>Design and conceptualized study; data interpretation; drafted the manuscript for intellectual content</td>
</tr>
</tbody>
</table>
The discerning eye of computer vision: can it measure Parkinson’s finger tap bradykinesia?

Authors: Stefan Williams, MBBS, BSc; Zhibin Zhao, BEng; Awais Hafeez, MS, BSc; David C Wong, PhD; Samuel D Relton, PhD; Hui Fang, PhD; Jane E Alty MD, MA, MBChir.

Stefan Williams, University of Leeds, Leeds Institute of Health Science, Leeds, UK

Zhibin Zhao, (1) University of Manchester, Division of Informatics, Imaging & Data Sciences, Manchester, UK; (2) Xi’an Jiaotong University, School of Mechanical Engineering, Xi’an, China.

Awais Hafeez, (1) University of Leeds, School of Mechanical Engineering, Leeds, UK; (2) University of Engineering and Technology Lahore, Department of Mechatronics and Control Engineering, Lahore, Pakistan

David C Wong, University of Manchester, Division of Informatics, Imaging & Data Sciences, Manchester, UK

Samuel D Relton, University of Leeds, Leeds Institute of Health Science, Leeds, UK

Hui Fang, Loughborough University, Department of Computer Science, Loughborough, UK

Jane E Alty, (1) University of Tasmania, Wicking Dementia Research & Education Centre, Hobart, Australia (2) Leeds Teaching Hospitals NHS Trust UK.

Word count: 3099
Number of Tables: 3
Number of Figures: 3
Number of Videos: 1

Corresponding Author:
Stefan Williams
Leeds Institute of Health Science, 10th Floor, Worsley Building
University of Leeds, Clarendon Way, Leeds, UK LS2 9LU
Phone: 0113 343 0839
umswi@leeds.ac.uk
ORCID: 0000-0002-4786-1356

Keywords:
Parkinson’s disease; Parkinsonism; Bradykinesia; Finger tapping; Computer Vision; Artificial Intelligence; DeepLabCut
ABSTRACT

Objective
The worldwide prevalence of Parkinson’s disease is increasing. There is urgent need for new tools to objectively measure the condition. Existing methods to record the cardinal motor feature of the condition, bradykinesia, using wearable sensors or smartphone apps have not reached large-scale, routine use. We evaluate new computer vision (artificial intelligence) technology, DeepLabCut, as a contactless method to quantify measures related to Parkinson’s bradykinesia from smartphone videos of finger tapping.

Methods
Standard smartphone video recordings of 133 hands performing finger tapping (39 idiopathic Parkinson’s patients and 30 controls) were tracked on a frame-by-frame basis with DeepLabCut. Objective computer measures of tapping speed, amplitude and rhythm were correlated with clinical ratings made by 22 movement disorder neurologists using the Modified Bradykinesia Rating Scale (MBRS) and Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS).

Results
DeepLabCut reliably tracked and measured finger tapping in standard smartphone video. Computer measures correlated well with clinical ratings of bradykinesia (Spearman coefficients): -0.74 speed, 0.66 amplitude, -0.65 rhythm for MBRS; -0.56 speed, 0.61 amplitude, -0.50 rhythm for MDS-UPDRS; 0.69 combined for MDS-UPDRS. All p < 0.001.

Conclusion
New computer vision software, DeepLabCut, can quantify three measures related to Parkinson’s bradykinesia from smartphone videos of finger tapping. Objective ‘contactless’ measures of standard clinical examinations were not previously possible with wearable sensors (accelerometers, gyroscopes, infrared markers). DeepLabCut requires only conventional video recording of clinical examination and is entirely ‘contactless’. This next generation technology holds potential for Parkinson’s and other neurological disorders with altered movements.
INTRODUCTION

The worldwide prevalence of Parkinson’s is increasing, with an estimated 10 million people already affected [1]. There is urgent need to find new tools to objectively measure the condition, to assist diagnosis, monitoring, and research metrics [2]. This is especially pressing in the era of precision medicine and with a global shortage of neurologists [3].

The cornerstone of Parkinson’s assessment is the clinician’s visual judgement of bradykinesia: one of the cardinal motor features of the condition, which is defined as slowness of movement and decrement in size or speed, or progressive hesitations/halts, as movements are continued [4]. There are several clinical examination methods to assess bradykinesia, but in a brief clinic room assessment, one of the most common is to observe the patient repeatedly tapping their index finger against thumb ‘as quickly and as big as possible’ [5]. However, clinicians’ visual interpretation is inherently subjective [6,7], with considerable inter-rater variability [5,9-11]. Previous reported methods to measure bradykinesia on finger tapping have required wearable equipment such as gyroscopes [9], electromagnetic sensors [12; 13], and infrared camera markers [14], or patient interaction with an app-based task [15], and this has hindered the pathway to large-scale, routine clinical use.

Recent developments in computer interpretation of video (computer vision), clear the pathway to ‘contactless’ methods of quantifying clinical assessments [16], with potential widespread global application because of the availability of inexpensive smartphone cameras [17]. DeepLabCut is recently published open source computer vision software, without a requirement to write new computer code, that has been demonstrated to track and measure mice and fruit fly body parts [18,19]. We apply it to clinical neurology for the first time and assess whether the technology can be used to track and measure bradykinesia in standard smartphone video of conventional finger tapping examination.

METHODS
Ethical review

The study was approved by the North of Scotland Research Ethics Committee, United Kingdom Health Research Authority (IRAS project ID 256116).

Participants and video recording

Video recordings of 138 hands (including left and right) were obtained from 39 patients with idiopathic Parkinson’s disease and 30 healthy controls, who gave written consent. One video was rejected because the hand moved outside the video frame, making 137 videos: 77 Parkinson’s hands and 60 control hands.

All Parkinson’s patients had previously been diagnosed by a movement disorder specialist neurologist at Leeds Teaching Hospitals NHS Trust United Kingdom, according to Movement Disorder Society clinical diagnostic criteria [4], and were subjectively and objectively in the ‘on’ state at the time of video recording (no medication was withheld prior to recording). We did not exclude patients with postural hand tremor or dystonia. Controls were recruited from the companions of patients, or hospital/university staff, and had no history of Parkinson’s disease or other neurological diagnosis.

Participants rested their elbow on a chair arm with the forearm lifted to a 45 degree angle. The hand was free to move as per the protocol of the Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) Item 3.4 Finger Tapping [5]. Only the hand/forearm was within the video frame. The distance from camera to hand was approximately 1m, but not tightly defined. Digits 1 and 2 were closest to the camera. No specific instructions were given for the position of digits 3 to 5, and participants were free to position these digits as they preferred, although the researcher gave a brief demonstration in which digits 3 to 5 were extended.
A smartphone, placed on a tripod, was used to record standard video (60 frames per second, 1920x1080 pixel resolution), with only ambient lighting. Participants were instructed to tap their index finger and thumb together “as quickly and as big as possible” for at least 10 seconds [5]. Each video was edited to an 11 second clip: 1 second prior to tapping onset and 10 seconds of finger tapping.

Clinical rating

22 consultant neurologists specialising in movement disorders (United Kingdom) were asked to rate 30 tapping videos each, selected at random from the set of 137 videos. The median number of raters per video was 5 (range 1 to 12, interquartile range 3 to 7). The raters were blinded to patient / healthy control status and each other’s scores. Each video was rated according to both the MDS-UPDRS Item 3.4 Finger Tapping [5], and the Modified Bradykinesia Rating Scale (MBRS) [8,9]. The MDS-UPDRS requires the rater to amalgamate judgments of finger tapping speed, amplitude, and rhythm into a single composite score (Table 1). In contrast, the MBRS is comprised of three separate scores for speed, amplitude, and rhythm (Table 2). The modal clinician rating for each video was used for correlation with computer measures (see below). For each possible pair of raters, Cohen Kappa was calculated based on all videos they had both seen. The resulting values had mean 0.28 (fair agreement) and standard deviation 0.28.

DeepLabCut video tracking

DeepLabCut tracks the geometrical configuration of multiple body parts in video, without a requirement to wear markers [18,20]. It uses transfer learning, with feature detectors based on deep neural networks that have been pretrained on ImageNet, a massive object recognition dataset [21]. This means that DeepLabCut is able to accurately recognise and track body parts with minimal training data [18].

The set of 137 finger tapping videos was processed by DeepLabCut. We localised (labelled) six distinct points on the tapping hand in 20 frames selected by k-means clustering from each 660 frame video:
thumb tip; index finger tip; thumb metacarpophalangeal (MCP) joint; index finger MCP joint; middle finger tip; dorsal wrist/proximal dorsal hand. The deep neural network architecture of DeepLabCut was then trained using these points to predict their localisation in the remaining 97% of (unlabelled) video frames (103000 training iterations, ResNet 50). This created video pixel coordinates for finger tip and thumb tip throughout each video.

The accuracy of DeepLabCut to track the hand localisation points (including finger tip and thumb tip) was assessed using the ‘evaluate network’ function within DeepLabCut. This function computes the Euclidean error between the manual labels and the ones predicted by DeepLabCut averaged over the hand locations and test images (mean absolute error, proportional to the average root mean square error) [20].

Signal processing

Video pixel coordinates for the labels produced by DeepLabCut were used to calculate the pixel distance between index finger tip and thumb tip. A Savitzky-Golay filter was applied to the resulting time series, which removed large, sudden transient label ‘jumps’ caused by DeepLabCut mislabelling (i.e. large-distance, physiologically impossible label movement across one pair of video frames). The pixel number distances were standardised across all videos, by using the maximum opening distance between finger and thumb tip and normalising this to a value of 1 (all values were divided by the maximal value in the corresponding time series).

Three features of the resulting finger tip to thumb tip distance time series were calculated to reflect the clinical features of finger tapping speed, amplitude and rhythm (Figure 1). A measure of speed was calculated as the mean rate of change of the normalised distance between finger and thumb tip over time. A measure of amplitude variability was calculated by dividing each time series into one-second windows, with maximal overlap, and then calculating the coefficient of variation of the mean difference between the maximum and minimum amplitudes in each window. A measure of rhythm regularity was
calculated by undertaking Fast Fourier Transform to find the distribution of frequencies within each finger tap time series, and then measuring the power of the dominant frequency peak added to the power of the frequencies 0.2 Hz either side of it. A more regular tapping rhythm concentrates power in a narrow frequency band, increasing the power of the dominant frequency peak (and its immediate neighbours), whereas a more irregular tapping rhythm leads to a more widely spread distribution of frequency bands, reducing the power of the dominant frequency peak (and its immediate neighbours).

Spearman correlation coefficients were calculated for the each of the three DeepLabCut measures vs the modal MBRS and MDS-UPDRS clinician ratings. In addition, we calculated the Spearman correlation coefficient of the three computer scores combined (the normalised arithmetic mean of the computer speed score, amplitude variability score and rhythm regularity score) with the MDS-UPDRS rating.

RESULTS

The participant and hand video characteristics are given in Table 3.

Tracking accuracy

DeepLabCut reliably tracked the finger tapping movements across 133 videoed clinical examinations. Figure 2 shows example frames from a video labelled by DeepLabCut. The mean absolute error of DeepLabCut labelling was 8.39 pixels, i.e. the average distance between manual (human) labels and those predicted by DeepLabCut was 8.39 pixels within a 1920x1080 pixel video frame. Video 1 shows examples of four videos labelled by DeepLabCut, including particularly challenging cases with tremor and dystonia also present.

Correlation with clinical bradykinesia ratings
The quantitative scores derived from DeepLabCut measurements correlated well with the group neurologists’ modal ratings of bradykinesia; Figure 3. The computer vision measures of tapping speed, amplitude variability and rhythm regularity had good correlations with the respective MBRS clinical ratings for speed (-0.70, p < 0.001), amplitude (0.65, p < 0.001) and rhythm (-0.61, p < 0.001). There were also good correlations between the computer measures of tapping speed (-0.66, p < 0.001), amplitude variability (0.56, p < 0.001) and rhythm regularity (-0.50, p < 0.001) and the MDS-UPDRS clinical rating. The MDS-UPDRS is a composite clinical rating, combining the separate components of bradykinesia, and the three computer measures combined also correlated with MDS-UPDRS (-0.69, p < 0.001).

DISCUSSION

We have shown that new computer vision software, DeepLabCut, can track finger tapping in a standard smartphone video recording of a conventional clinical examination, and quantify three measures related to Parkinson’s bradykinesia: speed, amplitude variation and rhythm regularity. We found good correlations between the objective computer measures of finger tapping and clinical ratings of bradykinesia by 22 movement disorder neurologists using two gold standard clinical scales (MBRS and MDS-UPDRS). Thus, smartphone video measures using DeepLabCut exhibit convergent validity with conventional clinician ratings.

These results are new and clinically significant. This is the first study to apply new open source DeepLabCut technology, developed for animal behaviour studies, to the field of human movement disorders, or indeed any aspect of clinical neurology. We demonstrate new objective measures related to bradykinesia by bringing together two ubiquitous ingredients: the smartphone video camera and visual artificial intelligence software. With the worldwide ownership of smartphones so high (e.g. 82% in the UK, 77% in the US) [17], there are no hardware barriers to further use of this approach.
Neurologists recognize that quantification of bradykinesia will improve Parkinson’s management through ‘precision medicine’ principles: it has the potential to aid early accurate diagnosis and remote monitoring, and provide fine grain regular measurement of clinical trial outcomes. Previous methods to quantify bradykinesia have focussed on wearable sensors measuring finger tapping [9,12-14,23] but it has been challenging to incorporate these into widespread use because equipment needs to be physically attached to the patient. Furthermore, the necessity for sensors to stay fixed on a specific body part has often required the exclusion of tremulous and dyskinetic patients [8,9,12,14,23]. In contrast, the DeepLabCut test presented here is a contactless method that simply requires a short 10 second video recording of the standard clinical examination. This means it can potentially be used in any setting and any patient.

The ‘real life’ versatility of DeepLabCut in ordinary clinical settings was evaluated in this study because no special efforts were made to optimize participant positioning or filming conditions. The tracking and measures were robust despite variations in smartphone camera distance, positioning of the hand, skin color, flexion / extension postures of digits 3 to 5, ambient lighting conditions, and inclusion of a small number of patients with tremor or dystonia during tapping. Video 1 illustrates this, with accurate labelling of finger and thumb tip despite superimposed tremor or dystonia. This is particularly encouraging for the relevance of DeepLabCut aiding assessment of Parkinson’s in conventional clinical settings.

Two further strengths of our study are the collection of clinical ratings from a large group of movement disorder neurologists, and the use of two rating scales: the MBRS and MDS-UPDRS, such that the clinical ratings were representative and reliable. Clinical rating scales are vulnerable to inter-rater variability but using modal values derived from a large number of blinded clinical raters, and two different gold-standard rating scales, provided a more robust ‘ground truth’ of clinical bradykinesia to evaluate the new test against. As such we were able to demonstrate that clinician and DeepLabCut measures correlate for both individual components of bradykinesia (MBRS) and an overall combination...
of those components (MDS-UPRDS). An interesting future application with a larger participant group
would be to examine the relative contribution of the three video measures to individual grades of
UPDRS bradykinesia.

Another advantage of the approach described here is that it is not limited by the physical constraints of
wearable or screen/keyboard tapping methods, and could be applied to detect and measure other motor
features of Parkinson’s in a contactless manner such as tremor, gait, or posture. Computer vision is
general has potential for application to these signs [24,25].

Only three previous studies have objectively measured finger tapping using contactless, standard video
analysis. One method extracted finger tapping information from 13 Parkinson's patients with advanced
disease [26], but it required inclusion of the patient's face in the video, limiting the practical utility. The
second report, from our own group, described a video method to measure movement using an optical
flow field [27] that interprets movements of the whole hand. The work here improves on this by
explicitly extracting the salient thumb and finger tip points, and validating in a larger cohort. A third
study of 60 Parkinson’s patients (no controls) found strong correlation between MDS-UPRDS finger
tap rating and video measures of tap interval (frequency) (r=0.91), frequency variation (r=0.82) and
amplitude (r=-0.94) but not amplitude variation (r=0.39) [28]. The ‘ground truth’ of the clinical
comparison was less robust, based on just two clinical raters, and there were no MBRS ratings;
furthermore, some images were blurred in the 24 frame per second videos. A major advantage of our
method in comparison is that DeepLabCut is open source software, that can be downloaded and used
without the need to write computer code, making it available now to neurologists with only limited
computing knowledge.

A well-studied, non-camera method to measure finger tap bradykinesia is the ‘Bradykinesia-Akinesia
Incoordination (BRAIN) Test’, which involves participants tapping a standard computer keyboard for
30 seconds [29]. This shares advantages with our method in that it involves no special equipment, and
could be employed without clinician-patient contact. However, it requires patient motivation to engage
with a keyboard tapping task, rather than a simple recording of existing standard clinical examination (and similar requirements for patient motivation apply to patient smartphone apps [15]). Furthermore, ‘BRAIN’ correlations with MDS-UPDRS finger tapping ratings of 0.44, 0.20, 0.33, 0.03 [29], are weaker than those we report (0.66, 0.56, 0.50, 0.69). One reason for this may be that a computer keyboard cannot record any measure of tap amplitude, and the speed measure is limited to tap frequency, so that some core aspects of bradykinesia cannot be captured by ‘BRAIN’.

The present study has some limitations. Like human vision, a simple camera lacks an absolute measure of distance, and our method cannot capture 3D movement. Rotation of the thumb and finger into a horizontal plane would falsely alter amplitude measurement. However, in practice such movement is rare during tapping, and our normalized measure of finger-thumb distance based on 2D maximum opening distance appeared to capture amplitude, speed and rhythm abnormalities well, given the good correlations with clinical ratings.

We have measured amplitude variation, but not amplitude decrement (the sequence effect). Previous approaches have measured decrement as a straight line linear regression [23], but decrement would also proportionately affect a measure of amplitude variation, such that the variations in decrement should be captured within our measure. Furthermore, there are several other previous reports of good correlation between non-decrement amplitude measures using sensors and clinician ratings of finger tap bradykinesia [9].

Relatively few videos were rated as grade 3 or 4 bradykinesia, and it is possible that a greater number of higher grade videos might change the strength of correlations. Even with accurate tracking, all forms of rating or measuring bradykinesia (including clinician) may potentially be confounded by comorbidities such as joint deformities, pain, dystonia etc. However, with a larger dataset, machine learning techniques are well-suited to separate such additional contributions from underlying bradykinesia. There was a degree of error in DeepLabCut labelling of finger and thumb tips, but this was minimal (mean of 8.39 pixels within a 1920x1080 pixel video frame) and unlikely to affect...
individual measures. We mitigated against this by applying a filter prior to feature extraction. Full flexion of digits 3 to 5 might improve labelling accuracy, but would also limit the amplitude of finger tapping [23], potentially masking more subtle deficits. The labelling error provided by DeepLabCut’s ‘evaluate network’ function only gives an accuracy figure for the entire set of videos but not for individual videos, meaning that we have not reported accuracy in subsets of videos, such as those with visible tremor. This study was exploratory, aimed at proof of concept that bradykinesia related measures could be derived from smartphone video. The protocol was thus simple, and repeat measures such as test-retest in controls, or ‘off’ and ‘on’ medication states, were not included, meaning that reliability and sensitivity to change have not been characterised. Future work will assess this.

Finally, we have demonstrated measurements that show convergent validity with clinical ratings, rather than a test to discriminate patients from controls. In our view, it makes little sense to pursue automated diagnostic classification using finger tapping video alone because clinical Parkinson’s diagnosis is by definition a broader assessment than isolated finger tapping performance [4].

This study has evaluated the clinical application of DeepLabCut in a medical condition, specifically a neurological disorder. The methodology described here provides a new objective measure of Parkinson’s bradykinesia in conventional clinical examination. DeepLabCut provides a contactless means for tracking and measuring bradykinesia in Parkinson’s, without any special equipment other than the ubiquitous smartphone. This research therefore provides a new tool to quantify bradykinesia. It could potentially be used to support diagnosis and monitoring of Parkinson’s, both clinically and as an outcome measure in research studies. Beyond Parkinson’s and other movement disorders, this same technology could be applied to many other neurological conditions that require measurement of movements or function; for example multiple sclerosis, neuropathies, epilepsy and stroke.

REFERENCES


**Conflict of Interest Statement**

The authors declare that they have no conflict of interest.

**FIGURE LEGENDS**
**Figure 1.** Illustration of the three parameters derived from the DeepLabCut finger tip and thumb tip coordinates to measure finger tapping speed, amplitude and rhythm.

**Figure 2.** Example video frames taken from smartphone video labelled by DeepLabCut, showing the six localization (labelling) points: thumb tip (dark blue), thumb metacarpophalangeal (MCP) joint (light blue), index finger tip (cyan), index finger MCP joint (yellow), middle finger tip (orange), dorsal wrist (red).

**Figure 3.** Computer measures of finger tapping speed, amplitude variation and rhythm, derived from DeepLabCut smartphone video tracking, correlate with clinician mode ratings for the same videos. Pearson correlation coefficients shown. MBRS: Modified Bradykinesia Rating Scale. MDS-UPDRS: Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale.

**TABLES**

**Table 1.** The Movement Disorder Society revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Item 3.4 (Finger Tapping).

<table>
<thead>
<tr>
<th>MDS-UPDRS Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: Normal</td>
<td>No problems</td>
</tr>
<tr>
<td>1: Slight</td>
<td>Any of the following: a) the regular rhythm is broken with one or two interruptions or hesitations of the tapping movement; b) slight slowing; c) the amplitude decrements near the end of the 10 taps.</td>
</tr>
<tr>
<td>2: Mild</td>
<td>Any of the following: a) 3 to 5 interruptions during tapping; b) mild slowing; c) the amplitude decrements midway in the 10-tap sequence.</td>
</tr>
<tr>
<td>3: Moderate</td>
<td>Any of the following: a) more than 5 interruptions during tapping or at least one longer arrest (freeze) in ongoing movement; b) moderate slowing; c) the amplitude decrements starting after the 1st tap.</td>
</tr>
<tr>
<td>4: Severe</td>
<td>Cannot or can only barely perform the task because of slowing, interruptions, or decrements.</td>
</tr>
</tbody>
</table>
Each hand is tested separately. The patient is instructed to tap the index finger on the thumb "as quickly and as big as possible". The first 10 taps are evaluated by the clinician [5].

Table 2. The Modified Bradykinesia Rating Scale (MBRS).

<table>
<thead>
<tr>
<th>Score</th>
<th>Speed</th>
<th>Amplitude</th>
<th>Rhythm</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal</td>
<td>Normal</td>
<td>Regular, no arrests or pauses in ongoing movement</td>
</tr>
<tr>
<td>1</td>
<td>Mild Slowing</td>
<td>Mild reduction in amplitude in later performance, most movements close to normal</td>
<td>Mild impairment, up to two brief arrests in the 10 seconds, none lasting &gt; 1 second</td>
</tr>
<tr>
<td>2</td>
<td>Moderate Slowing</td>
<td>Moderate, reduction in amplitude visible early in performance but continues to maintain 50% amplitude through most of the tasks</td>
<td>Moderate, 3 to 4 arrests in 10 seconds; OR 1 or 2 lasting &gt; 1 second</td>
</tr>
<tr>
<td>3</td>
<td>Severe slowing</td>
<td>Severe, less than 50% amplitude through most of the task</td>
<td>Severe, 5 or more arrests/10 seconds; OR more than 2 lasting &gt; 1 second</td>
</tr>
<tr>
<td>4</td>
<td>Can barely perform the task</td>
<td>Can barely perform the task</td>
<td>Can barely perform the task</td>
</tr>
</tbody>
</table>

Each hand is tested separately. The patient is instructed to tap the index finger on the thumb "as quickly and as big as possible". The first 10 seconds of tapping are evaluated by the clinician to give three MBRS ratings (one for each of speed, amplitude, rhythm) [6,7].

Table 3. Study participant (hand video) characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Patients</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Std. Dev.) yrs</td>
<td>68 (9.6)</td>
<td>59 (19.4)</td>
</tr>
<tr>
<td>Male/Female</td>
<td>47/26</td>
<td>22/38</td>
</tr>
<tr>
<td>Median years since diagnosis</td>
<td>4</td>
<td>n/a</td>
</tr>
<tr>
<td>Median H&amp;Y [IQR]</td>
<td>2 [1,3]</td>
<td>n/a</td>
</tr>
<tr>
<td>H&amp;Y = 1</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 1.5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 2</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 2.5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 3</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 4</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
**H&Y = 5**

<table>
<thead>
<tr>
<th>Median MDS-UPDRS [IQR]</th>
<th>2 [1,3]</th>
<th>1 [0,1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDS-UPDRS = 0</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td>MDS-UPDRS = 1</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>MDS-UPDRS = 2</td>
<td>19</td>
<td>12</td>
</tr>
<tr>
<td>MDS-UPDRS = 3</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>MDS-UPDRS = 4</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Visible tremor in video</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Visible dystonia in video</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Data is split by Parkinson's hands (n=73) and control hands (n=60). H&Y: modified Hoehn and Yahr scale [22]. MDS-UPDRS: Movement Disorder Society revision of the Unified Parkinson's Disease Rating Scale, Item 3.4 (Finger Tapping). IQR: Interquartile Range.

**Acknowledgements**

We would like to thank the following UK movement disorder neurologists for providing clinical rating of videos: Dr Jane Alty, Dr Sundus Alusi, Dr Michael Bonello, Dr Stephen Butterworth, Dr Philip Buttery, Dr Camille Carroll, Dr Adam Cassidy, Dr Jeremy Cosgrove, Dr Richard Ellis, Dr Jonathan Evans, Dr Paul Goldsmith, Dr Donald Grosset, Dr Christopher Kobylecki, Dr Alistair Lansbury, Dr Peter Moore, Dr Rachel Newby, Dr Edward Newman, Dr Gillian Sare, Dr Monty Silverdale, Dr Naomi Warren, Dr Louise Wiblin, Dr Caroline Williams-Gray.

We would also like to thank Dr Parisa Patel and Dr Paschal O’Gorman for data collection.

**Author contributions**

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Role</th>
<th>Contribution</th>
</tr>
</thead>
</table>

C:\pdfbuilder\temp\Revision(marked)\Computer_Vision_Parkinsons_Bradykinesia.docx
<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stefan Williams, MBBS, BSc</strong></td>
<td>University of Leeds, UK</td>
<td>Design and conceptualized study; data acquisition and interpretation; drafted the manuscript for intellectual content</td>
</tr>
<tr>
<td><strong>Zhibin Zhao, BEng</strong></td>
<td>(1) University of Manchester, UK; (2) Xi’an Jiaotong University, China</td>
<td>Analyzed the data (major role)</td>
</tr>
<tr>
<td><strong>Awais Hafeez, MA, BSc</strong></td>
<td>(1) University of Leeds, UK; (2) University of Engineering and Technology Lahore, Pakistan</td>
<td>Design and conceptualized the study; major role in acquisition of data</td>
</tr>
<tr>
<td><strong>David C Wong, PhD</strong></td>
<td>University of Manchester, UK</td>
<td>Design and conceptualized study; analyzed the data; data interpretation; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td><strong>Samuel D Relton, PhD</strong></td>
<td>University of Leeds, UK</td>
<td>Design and conceptualized study; data interpretation; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td><strong>Hui Fang, PhD</strong></td>
<td>Loughborough University, UK</td>
<td>Design and conceptualized study (major role)</td>
</tr>
<tr>
<td><strong>Jane E Alty, MD, MA, MBBChir</strong></td>
<td>(1) University of Tasmania, Australia; (2) Leeds Teaching Hospitals NHS Trust, UK</td>
<td>Design and conceptualized study; data interpretation; drafted the manuscript for intellectual content</td>
</tr>
</tbody>
</table>
The discerning eye of computer vision: can it measure Parkinson’s finger tap bradykinesia?

Authors: Stefan Williams, MBBS, BSc; Zhibin Zhao, BEng; Awais Hafeez, MS, BSc; David C Wong, PhD; Samuel D Relton, PhD; Hui Fang, PhD; Jane E Alty MD, MA, MBBChir.

Stefan Williams, University of Leeds, Leeds Institute of Health Science, Leeds, UK

Zhibin Zhao, (1) University of Manchester, Division of Informatics, Imaging & Data Sciences, Manchester, UK; (2) Xi’an Jiatong University, School of Mechanical Engineering, Xi’an, China.

Awais Hafeez, (1) University of Leeds, School of Mechanical Engineering, Leeds, UK; (2) University of Engineering and Technology Lahore, Department of Mechatronics and Control Engineering, Lahore, Pakistan

David C Wong, University of Manchester, Division of Informatics, Imaging & Data Sciences, Manchester, UK

Samuel D Relton, University of Leeds, Leeds Institute of Health Science, Leeds, UK

Hui Fang, Loughborough University, Department of Computer Science, Loughborough, UK

Jane E Alty, (1) University of Tasmania, Wicking Dementia Research & Education Centre, Hobart, Australia (2) Leeds Teaching Hospitals NHS Trust UK.

Word count: 3101
Number of Tables: 3
Number of Figures: 3
Number of Videos: 1

Corresponding Author:
Stefan Williams
Leeds Institute of Health Science, 10th Floor, Worsley Building
University of Leeds, Clarendon Way, Leeds, UK LS2 9LU
Phone: 0113 343 0839
umswi@leeds.ac.uk
ORCID: 0000-0002-4786-1356

Keywords:
Parkinson’s disease; Parkinsonism; Bradykinesia; Finger tapping; Computer Vision; Artificial Intelligence; DeepLabCut
ABSTRACT

Objective

The worldwide prevalence of Parkinson’s disease is increasing. There is urgent need for new tools to objectively measure the condition. Existing methods to record the cardinal motor feature of the condition, bradykinesia, using wearable sensors or smartphone apps have not reached large-scale, routine use. We evaluate new computer vision (artificial intelligence) technology, DeepLabCut, as a contactless method to quantify measures related to Parkinson’s bradykinesia from smartphone videos of finger tapping.

Methods

Standard smartphone video recordings of 133 hands performing finger tapping (39 idiopathic Parkinson’s patients and 30 controls) were tracked on a frame-by-frame basis with DeepLabCut. Objective computer measures of tapping speed, amplitude and rhythm were correlated with clinical ratings made by 22 movement disorder neurologists using the Modified Bradykinesia Rating Scale (MBRS) and Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS).

Results

DeepLabCut reliably tracked and measured finger tapping in standard smartphone video. Computer measures correlated well with clinical ratings of bradykinesia (Spearman coefficients): \(-0.74\) speed, \(0.66\) amplitude, \(-0.65\) rhythm for MBRS; \(-0.56\) speed, \(0.61\) amplitude, \(-0.50\) rhythm for MDS-UPDRS; \(0.69\) combined for MDS-UPDRS. All \(p < 0.001\).

Conclusion

New computer vision software, DeepLabCut, can quantify three measures related to Parkinson’s bradykinesia from smartphone videos of finger tapping. Objective ‘contactless’ measures of standard clinical examinations were not previously possible with wearable sensors (accelerometers, gyroscopes, infrared markers). DeepLabCut requires only conventional video recording of clinical examination and is entirely ‘contactless’. This next generation technology holds potential for Parkinson’s and other neurological disorders with altered movements.
INTRODUCTION

The worldwide prevalence of Parkinson’s is increasing, with an estimated 10 million people already affected [1]. There is urgent need to find new tools to objectively measure the condition, to assist diagnosis, monitoring, and research metrics [2]. This is especially pressing in the era of precision medicine and with a global shortage of neurologists [3].

The cornerstone of Parkinson’s assessment is the clinician’s visual judgement of bradykinesia: one of the cardinal motor features of the condition, which is defined as slowness of movement and decrement in size or speed, or progressive hesitations/halts, as movements are continued [4]. There are several clinical examination methods to assess bradykinesia, but in a brief clinic room assessment, one of the most common is to observe the patient repeatedly tapping their index finger against thumb ‘as quickly and as big as possible’ [5]. However, clinicians’ visual interpretation is inherently subjective [6,7], with considerable inter-rater variability [5,9-11]. Previous reported methods to measure bradykinesia on finger tapping have required wearable equipment such as gyroscopes [9], electromagnetic sensors [12; 13], and infrared camera markers [14], or patient interaction with an app-based task [15], and this has hindered the pathway to large-scale, routine clinical use.

Recent developments in computer interpretation of video (computer vision), clear the pathway to ‘contactless’ methods of quantifying clinical assessments [16], with potential widespread global application because of the availability of inexpensive smartphone cameras [17]. DeepLabCut is recently published open source computer vision software, without a requirement to write new computer code, that has been demonstrated to track and measure mice and fruit fly body parts [18,19]. We apply it to clinical neurology for the first time and assess whether the technology can be used to track and measure bradykinesia in standard smartphone video of conventional finger tapping examination.

METHODS
Ethical review

The study was approved by the North of Scotland Research Ethics Committee, United Kingdom Health Research Authority (IRAS project ID 256116).

Participants and video recording

Video recordings of 138 hands (including left and right) were obtained from 39 patients with idiopathic Parkinson’s disease and 30 healthy controls, who gave written consent. One video was rejected because the hand moved outside the video frame, making 137 videos: 77 Parkinson’s hands and 60 control hands.

All Parkinson’s patients had previously been diagnosed by a movement disorder specialist neurologist at Leeds Teaching Hospitals NHS Trust United Kingdom, according to Movement Disorder Society clinical diagnostic criteria [4], and were subjectively and objectively in the ‘on’ state at the time of video recording (no medication was withheld prior to recording). We did not exclude patients with postural hand tremor or dystonia. Controls were recruited from the companions of patients, or hospital/university staff, and had no history of Parkinson’s disease or other neurological diagnosis.

Participants rested their elbow on a chair arm with the forearm lifted to a 45 degree angle. The hand was free to move as per the protocol of the Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) Item 3.4 Finger Tapping [5]. Only the hand/forearm was within the video frame. The distance from camera to hand was approximately 1m, but not tightly defined. Digits 1 and 2 were closest to the camera. No specific instructions were given for the position of digits 3 to 5, and participants were free to position these digits as they preferred, although the researcher gave a brief demonstration in which digits 3 to 5 were extended.
A smartphone, placed on a tripod, was used to record standard video (60 frames per second, 1920x1080 pixel resolution), with only ambient lighting. Participants were instructed to tap their index finger and thumb together “as quickly and as big as possible” for at least 10 seconds [5]. Each video was edited to an 11 second clip: 1 second prior to tapping onset and 10 seconds of finger tapping.

Clinical rating

22 consultant neurologists specialising in movement disorders (United Kingdom) were asked to rate 30 tapping videos each, selected at random from the set of 137 videos. The median number of raters per video was 5 (range 1 to 12, interquartile range 3 to 7). The raters were blinded to patient / healthy control status and each other’s scores. Each video was rated according to both the MDS-UPDRS Item 3.4 Finger Tapping [5], and the Modified Bradykinesia Rating Scale (MBRS) [8,9]. The MDS-UPDRS requires the rater to amalgamate judgments of finger tapping speed, amplitude, and rhythm into a single composite score (Table 1). In contrast, the MBRS is comprised of three separate scores for speed, amplitude, and rhythm (Table 2). The modal clinician rating for each video was used for correlation with computer measures (see below). For each possible pair of raters, Cohen Kappa was calculated based on all videos they had both seen. The resulting values had mean 0.28 (fair agreement) and standard deviation 0.28.

DeepLabCut video tracking

DeepLabCut tracks the geometrical configuration of multiple body parts in video, without a requirement to wear markers [18,20]. It uses transfer learning, with feature detectors based on deep neural networks that have been pretrained on ImageNet, a massive object recognition dataset [21]. This means that DeepLabCut is able to accurately recognise and track body parts with minimal training data [18].

The set of 137 finger tapping videos was processed by DeepLabCut. We localised (labelled) six distinct points on the tapping hand in 20 frames selected by k-means clustering from each 660 frame video:
thumb tip; index finger tip; thumb metacarpophalangeal (MCP) joint; index finger MCP joint; middle finger tip; dorsal wrist/proximal dorsal hand. The deep neural network architecture of DeepLabCut was then trained using these points to predict their localisation in the remaining 97% of (unlabelled) video frames (1030000 training iterations, ResNet 50). This created video pixel coordinates for finger tip and thumb tip throughout each video.

The accuracy of DeepLabCut to track the hand localisation points (including finger tip and thumb tip) was assessed using the ‘evaluate network’ function within DeepLabCut. This function computes the Euclidean error between the manual labels and the ones predicted by DeepLabCut averaged over the hand locations and test images (mean absolute error, proportional to the average root mean square error) [20].

Signal processing

Video pixel coordinates for the labels produced by DeepLabCut were used to calculate the pixel distance between index finger tip and thumb tip. A Savitzky-Golay filter was applied to the resulting time series, which removed large, sudden transient label ‘jumps’ caused by DeepLabCut mislabelling (i.e. large-distance, physiologically impossible label movement across one pair of video frames). The pixel number distances were standardised across all videos, by using the maximum opening distance between finger and thumb tip and normalising this to a value of 1 (all values were divided by the maximal value in the corresponding time series).

Three features of the resulting finger tip to thumb tip distance time series were calculated to reflect the clinical features of finger tapping speed, amplitude and rhythm (Figure 1). A measure of speed was calculated as the mean rate of change of the normalised distance between finger and thumb tip over time. A measure of amplitude variability was calculated by dividing each time series into one-second windows, with maximal overlap, and then calculating the coefficient of variation of the mean difference between the maximum and minimum amplitudes in each window. A measure of rhythm regularity was
calculated by undertaking Fast Fourier Transform to find the distribution of frequencies within each finger tap time series, and then measuring the power of the dominant frequency peak added to the power of the frequencies 0.2 Hz either side of it. A more regular tapping rhythm concentrates power in a narrow frequency band, increasing the power of the dominant frequency peak (and its immediate neighbours), whereas a more irregular tapping rhythm leads to a more widely spread distribution of frequency bands, reducing the power of the dominant frequency peak (and its immediate neighbours).

Spearman correlation coefficients were calculated for the each of the three DeepLabCut measures vs the modal MBRS and MDS-UPDRS clinician ratings. In addition, we calculated the Spearman correlation coefficient of the three computer scores combined (the normalised arithmetic mean of the computer speed score, amplitude variability score and rhythm regularity score) with the MDS-UPDRS rating.

RESULTS

The participant and hand video characteristics are given in Table 3.

Tracking accuracy

DeepLabCut reliably tracked the finger tapping movements across 133 videoed clinical examinations. Figure 2 shows example frames from a video labelled by DeepLabCut. The mean absolute error of DeepLabCut labelling was 8.39 pixels, i.e. the average distance between manual (human) labels and those predicted by DeepLabCut was 8.39 pixels within a 1920x1080 pixel video frame. Video 1 shows examples of four videos labelled by DeepLabCut, including particularly challenging cases with tremor and dystonia also present.

Correlation with clinical bradykinesia ratings
The quantitative scores derived from DeepLabCut measurements correlated well with the group neurologists’ modal ratings of bradykinesia; Figure 3. The computer vision measures of tapping speed, amplitude variability and rhythm regularity had good correlations with the respective MBRS clinical ratings for speed (−0.70, p < 0.001), amplitude (0.65, p < 0.001) and rhythm (−0.61, p < 0.001). There were also good correlations between the computer measures of tapping speed (−0.66, p < 0.001), amplitude variability (0.56, p < 0.001) and rhythm regularity (−0.50, p < 0.001) and the MDS-UPDRS clinical rating. The MDS-UPDRS is a composite clinical rating, combining the separate components of bradykinesia, and the three computer measures combined also correlated with MDS-UPDRS (−0.69, p < 0.001).

DISCUSSION

We have shown that new computer vision software, DeepLabCut, can track finger tapping in a standard smartphone video recording of a conventional clinical examination, and quantify three measures related to Parkinson’s bradykinesia: speed, amplitude variation and rhythm regularity. We found good correlations between the objective computer measures of finger tapping and clinical ratings of bradykinesia by 22 movement disorder neurologists using two gold standard clinical scales (MBRS and MDS-UPDRS). Thus, smartphone video measures using DeepLabCut exhibit convergent validity with conventional clinician ratings.

These results are new and clinically significant. This is the first study to apply new open source DeepLabCut technology, developed for animal behaviour studies, to the field of human movement disorders, or indeed any aspect of clinical neurology. We demonstrate new objective measures related to bradykinesia by bringing together two ubiquitous ingredients: the smartphone video camera and visual artificial intelligence software. With the worldwide ownership of smartphones so high (e.g. 82% in the UK, 77% in the US) [17], there are no hardware barriers to further use of this approach.
Neurologists recognize that quantification of bradykinesia will improve Parkinson’s management through ‘precision medicine’ principles: it has the potential to aid early accurate diagnosis and remote monitoring, and provide fine grain regular measurement of clinical trial outcomes. Previous methods to quantify bradykinesia have focussed on wearable sensors measuring finger tapping [9,12-14,23] but it has been challenging to incorporate these into widespread use because equipment needs to be physically attached to the patient. Furthermore, the necessity for sensors to stay fixed on a specific body part has often required the exclusion of tremulous and dyskinetic patients [8,9,12,14,23]. In contrast, the DeepLabCut test presented here is a contactless method that simply requires a short 10 second video recording of the standard clinical examination. This means it can potentially be used in any setting and any patient.

The ‘real life’ versatility of DeepLabCut in ordinary clinical settings was evaluated in this study because no special efforts were made to optimize participant positioning or filming conditions. The tracking and measures were robust despite variations in smartphone camera distance, positioning of the hand, skin color, flexion / extension postures of digits 3 to 5, ambient lighting conditions, and inclusion of a small number of patients with tremor or dystonia during tapping. Video 1 illustrates this, with accurate labelling of finger and thumb tip despite superimposed tremor or dystonia. This is particularly encouraging for the relevance of DeepLabCut aiding assessment of Parkinson’s in conventional clinical settings.

Two further strengths of our study are the collection of clinical ratings from a large group of movement disorder neurologists, and the use of two rating scales: the MBRS and MDS-UPDRS, such that the clinical ratings were representative and reliable. Clinical rating scales are vulnerable to inter-rater variability but using modal values derived from a large number of blinded clinical raters, and two different gold-standard rating scales, provided a more robust ‘ground truth’ of clinical bradykinesia to evaluate the new test against. As such we were able to demonstrate that clinician and DeepLabCut measures correlate for both individual components of bradykinesia (MBRS) and an overall combination.
of those components (MDS-UPRDS). An interesting future application with a larger participant group would be to examine the relative contribution of the three video measures to individual grades of UPDRS bradykinesia.

Another advantage of the approach described here is that it is not limited by the physical constraints of wearable or screen/keyboard tapping methods, and could be applied to detect and measure other motor features of Parkinson’s in a contactless manner such as tremor, gait, or posture. Computer vision in general has potential for application to these signs [24,25].

Only three previous studies have objectively measured finger tapping using contactless, standard video analysis. One method extracted finger tapping information from 13 Parkinson's patients with advanced disease [26], but it required inclusion of the patient's face in the video, limiting the practical utility. The second report, from our own group, described a video method to measure movement using an optical flow field [27] that interprets movements of the whole hand. The work here improves on this by explicitly extracting the salient thumb and finger tip points, and validating in a larger cohort. A third study of 60 Parkinson’s patients (no controls) found strong correlation between MDS-UPRDS finger tap rating and video measures of tap interval (frequency) (r=0.91), frequency variation (r=0.82) and amplitude (r=-0.94) but not amplitude variation (r=0.39) [28]. The ‘ground truth’ of the clinical comparison was less robust, based on just two clinical raters, and there were no MBRS ratings; furthermore, some images were blurred in the 24 frame per second videos. A major advantage of our method in comparison is that DeepLabCut is open source software, that can be downloaded and used without the need to write computer code, making it available now to neurologists with only limited computing knowledge.

A well-studied, non-camera method to measure finger tap bradykinesia is the ‘Bradykinesia-Akinesia Incoordination (BRAIN) Test’, which involves participants tapping a standard computer keyboard for 30 seconds [29]. This shares advantages with our method in that it involves no special equipment, and could be employed without clinician-patient contact. However, it requires patient motivation to engage
with a keyboard tapping task, rather than a simple recording of existing standard clinical examination (and similar requirements for patient motivation apply to patient smartphone apps [15]). Furthermore, ‘BRAIN’ correlations with MDS-UPDRS finger tapping ratings of 0.44, 0.20, 0.33, 0.03 [29], are weaker than those we report (0.66, 0.56, 0.50, 0.69). One reason for this may be that a computer keyboard cannot record any measure of tap amplitude, and the speed measure is limited to tap frequency, so that some core aspects of bradykinesia cannot be captured by ‘BRAIN’.

The present study has some limitations. Like human vision, a simple camera lacks an absolute measure of distance, and our method cannot capture 3D movement. Rotation of the thumb and finger into a horizontal plane would falsely alter amplitude measurement. However, in practice such movement is rare during tapping, and our normalized measure of finger-thumb distance based on 2D maximum opening distance appeared to capture amplitude, speed and rhythm abnormalities well, given the good correlations with clinical ratings.

We have measured amplitude variation, but not amplitude decrement (the sequence effect). Previous approaches have measured decrement as a straight line linear regression [23], but decrement would also proportionately affect a measure of amplitude variation, such that the variations in decrement should be captured within our measure. Furthermore, there are several other previous reports of good correlation between non-decrement amplitude measures using sensors and clinician ratings of finger tap bradykinesia [9].

Relatively few videos were rated as grade 3 or 4 bradykinesia, and it is possible that a greater number of higher grade videos might change the strength of correlations. Even with accurate tracking, all forms of rating or measuring bradykinesia (including clinician) may potentially be confounded by comorbidities such as joint deformities, pain, dystonia etc. However, with a larger dataset, machine learning techniques are well-suited to separate such additional contributions from underlying bradykinesia. There was a degree of error in DeepLabCut labelling of finger and thumb tips, but this was minimal (mean of 8.39 pixels within a 1920x1080 pixel video frame) and unlikely to affect
individual measures. We mitigated against this by applying a filter prior to feature extraction. Full flexion of digits 3 to 5 might improve labelling accuracy, but would also limit the amplitude of finger tapping [23], potentially masking more subtle deficits. The labelling error provided by DeepLabCut’s ‘evaluate network’ function only gives an accuracy figure for the entire set of videos but not for individual videos, meaning that we have not reported accuracy in subsets of videos, such as those with visible tremor. This study was exploratory, aimed at proof of concept that bradykinesia related measures could be derived from smartphone video. The protocol was thus relatively simple. The test-retest reliability of this technique has not been addressed in the study but we hope to do so in future. Similarly, we did not measure ‘off’ and ‘on’ medication states (sensitivity to change).

Finally, we have demonstrated measurements that show convergent validity with clinical ratings, rather than a test to discriminate patients from controls. In our view, it makes little sense to pursue automated diagnostic classification using finger tapping video alone because clinical Parkinson’s diagnosis is by definition a broader assessment than isolated finger tapping performance [4].

This study has evaluated the clinical application of DeepLabCut in a medical condition, specifically a neurological disorder. The methodology described here provides a new objective measure of Parkinson’s bradykinesia in conventional clinical examination. DeepLabCut provides a contactless means for tracking and measuring bradykinesia in Parkinson’s, without any special equipment other than the ubiquitous smartphone. This research therefore provides a new tool to quantify bradykinesia. It could potentially be used to support diagnosis and monitoring of Parkinson’s, both clinically and as an outcome measure in research studies. Beyond Parkinson’s and other movement disorders, this same technology could be applied to many other neurological conditions that require measurement of movements or function; for example multiple sclerosis, neuropathies, epilepsy and stroke.

REFERENCES


**Conflict of Interest Statement**

The authors declare that they have no conflict of interest.

**FIGURE LEGENDS**
**Figure 1.** Illustration of the three parameters derived from the DeepLabCut finger tip and thumb tip coordinates to measure finger tapping speed, amplitude and rhythm.

**Figure 2.** Example video frames taken from smartphone video labelled by DeepLabCut, showing the six localization (labelling) points: thumb tip (dark blue), thumb metacarpophalangeal (MCP) joint (light blue), index finger tip (cyan), index finger MCP joint (yellow), middle finger tip (orange), dorsal wrist (red).

**Figure 3.** Computer measures of finger tapping speed, amplitude variation and rhythm, derived from DeepLabCut smartphone video tracking, correlate with clinician mode ratings for the same videos. Pearson correlation coefficients shown. MBRS: Modified Bradykinesia Rating Scale. MDS-UPDRS: Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale.

**TABLES**

**Table 1.** The Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) Item 3.4 (Finger Tapping).

<table>
<thead>
<tr>
<th>MDS-UPDRS Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: Normal</td>
<td>No problems</td>
</tr>
<tr>
<td>1: Slight</td>
<td>Any of the following: a) the regular rhythm is broken with one or two interruptions or hesitations of the tapping movement; b) slight slowing; c) the amplitude decrements near the end of the 10 taps.</td>
</tr>
<tr>
<td>2: Mild</td>
<td>Any of the following: a) 3 to 5 interruptions during tapping; b) mild slowing; c) the amplitude decrements midway in the 10-tap sequence.</td>
</tr>
<tr>
<td>3: Moderate</td>
<td>Any of the following: a) more than 5 interruptions during tapping or at least one longer arrest (freeze) in ongoing movement; b) moderate slowing; c) the amplitude decrements starting after the 1st tap.</td>
</tr>
<tr>
<td>4: Severe</td>
<td>Cannot or can only barely perform the task because of slowing, interruptions, or decrements.</td>
</tr>
</tbody>
</table>
Each hand is tested separately. The patient is instructed to tap the index finger on the thumb “as quickly and as big as possible”. The first 10 taps are evaluated by the clinician [5].

Table 2. The Modified Bradykinesia Rating Scale (MBRS).

<table>
<thead>
<tr>
<th>Score</th>
<th>Speed</th>
<th>Amplitude</th>
<th>Rhythm</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal</td>
<td>Normal</td>
<td>Regular, no arrests or pauses in ongoing movement</td>
</tr>
<tr>
<td>1</td>
<td>Mild Slowing</td>
<td>Mild reduction in amplitude in later performance, most movements close to normal</td>
<td>Mild impairment, up to two brief arrests in the 10 seconds, none lasting &gt; 1 second</td>
</tr>
<tr>
<td>2</td>
<td>Moderate Slowing</td>
<td>Moderate, reduction in amplitude visible early in performance but continues to maintain 50% amplitude through most of the tasks</td>
<td>Moderate, 3 to 4 arrests in 10 seconds; OR 1 or 2 lasting &gt; 1 second</td>
</tr>
<tr>
<td>3</td>
<td>Severe slowing</td>
<td>Severe, less than 50% amplitude through most of the task</td>
<td>Severe, 5 or more arrests/10 seconds; OR more than 2 lasting &gt; 1 second</td>
</tr>
<tr>
<td>4</td>
<td>Can barely perform the task</td>
<td>Can barely perform the task</td>
<td>Can barely perform the task</td>
</tr>
</tbody>
</table>

Each hand is tested separately. The patient is instructed to tap the index finger on the thumb “as quickly and as big as possible”. The first 10 seconds of tapping are evaluated by the clinician to give three MBRS ratings (one for each of speed, amplitude, rhythm) [6,7].

Table 3. Study participant (hand video) characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Patients</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Std. Dev.) yrs</td>
<td>68 (9.6)</td>
<td>59 (19.4)</td>
</tr>
<tr>
<td>Male/Female</td>
<td>47/26</td>
<td>22/38</td>
</tr>
<tr>
<td>Median years since diagnosis</td>
<td>4</td>
<td>n/a</td>
</tr>
<tr>
<td>Median H&amp;Y [IQR]</td>
<td>2 [1,3]</td>
<td>n/a</td>
</tr>
</tbody>
</table>

C:\pdfbuilder\temp\Second_Revision(CLEAN)ComputerVisionParkinsonsBradykinesia.docx
<table>
<thead>
<tr>
<th>H&amp;Y = 5</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median MDS-UPDRS (IQR)</td>
<td>2 [1,3]</td>
</tr>
<tr>
<td>MDS-UPDRS = 0</td>
<td>9</td>
</tr>
<tr>
<td>MDS-UPDRS = 1</td>
<td>20</td>
</tr>
<tr>
<td>MDS-UPDRS = 2</td>
<td>19</td>
</tr>
<tr>
<td>MDS-UPDRS = 3</td>
<td>20</td>
</tr>
<tr>
<td>MDS-UPDRS = 4</td>
<td>5</td>
</tr>
<tr>
<td>Visible tremor in video</td>
<td>11</td>
</tr>
<tr>
<td>Visible dystonia in video</td>
<td>2</td>
</tr>
</tbody>
</table>

Data is split by Parkinson’s hands (n=73) and control hands (n=60). H&Y: modified Hoehn and Yahr scale [22]. MDS-UPDRS: Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale, Item 3.4 (Finger Tapping). IQR: Interquartile Range.

Acknowledgements

We would like to thank the following UK movement disorder neurologists for providing clinical rating of videos: Dr Jane Alty, Dr Sundus Alusi, Dr Michael Bonello, Dr Stephen Butterworth, Dr Philip Buttery, Dr Camille Carroll, Dr Adam Cassidy, Dr Jeremy Cosgrove, Dr Richard Ellis, Dr Jonathan Evans, Dr Paul Goldsmith, Dr Donald Grosset, Dr Christopher Kobylecki, Dr Alistair Lansbury, Dr Peter Moore, Dr Rachel Newby, Dr Edward Newman, Dr Gillian Sare, Dr Monty Silverdale, Dr Naomi Warren, Dr Louise Wiblin, Dr Caroline Williams-Gray.

We would also like to thank Dr Parisa Patel and Dr Paschal O’Gorman for data collection.

Author contributions

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Role</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C:\pdf\temp\Second_Revision(CLEAN)ComputerVisionParkinsonsBradykinesia.docx
<table>
<thead>
<tr>
<th><strong>Stefan Williams,</strong> MBBS, BSc</th>
<th>University of Leeds, UK</th>
<th>Author</th>
<th>Design and conceptualized study; data acquisition and interpretation; drafted the manuscript for intellectual content</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Zhibin Zhao, BEng</strong></td>
<td>(1) University of Manchester, UK; (2) Xi’an Jiatong University, China</td>
<td>Author</td>
<td>Analyzed the data (major role)</td>
</tr>
<tr>
<td><strong>Awais Hafeez, MA, BSc</strong></td>
<td>(1) University of Leeds, UK; (2) University of Engineering and Technology Lahore, Pakistan</td>
<td>Author</td>
<td>Design and conceptualized the study; major role in acquisition of data</td>
</tr>
<tr>
<td><strong>David C Wong, PhD</strong></td>
<td>University of Manchester, UK</td>
<td>Author</td>
<td>Design and conceptualized study; analyzed the data; data interpretation; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td><strong>Samuel D Relton, PhD</strong></td>
<td>University of Leeds, UK</td>
<td>Author</td>
<td>Design and conceptualized study; data interpretation; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td><strong>Hui Fang, PhD</strong></td>
<td>Loughborough University, UK</td>
<td>Author</td>
<td>Design and conceptualized study (major role)</td>
</tr>
<tr>
<td><strong>Jane E Alty, MD, MA, MBBChir</strong></td>
<td>(1) University of Tasmania, Australia; (2) Leeds Teaching Hospitals NHS Trust, UK</td>
<td>Author</td>
<td>Design and conceptualized study; data interpretation; drafted the manuscript for intellectual content</td>
</tr>
</tbody>
</table>
The discerning eye of computer vision: can it measure Parkinson’s finger tap bradykinesia?

Authors: Stefan Williams, MBBS, BSc; Zhibin Zhao, BEng; Awais Hafeez, MS, BSc; David C Wong, PhD; Samuel D Relton, PhD; Hui Fang, PhD; Jane E Alty MD, MA, MBChir.

Stefan Williams, University of Leeds, Leeds Institute of Health Science, Leeds, UK

Zhibin Zhao, (1) University of Manchester, Division of Informatics, Imaging & Data Sciences, Manchester, UK; (2) Xi’an Jiaotong University, School of Mechanical Engineering, Xi’an, China.

Awais Hafeez, (1) University of Leeds, School of Mechanical Engineering, Leeds, UK; (2) University of Engineering and Technology Lahore, Department of Mechatronics and Control Engineering, Lahore, Pakistan

David C Wong, University of Manchester, Division of Informatics, Imaging & Data Sciences, Manchester, UK

Samuel D Relton, University of Leeds, Leeds Institute of Health Science, Leeds, UK

Hui Fang, Loughborough University, Department of Computer Science, Loughborough, UK

Jane E Alty, (1) University of Tasmania, Wicking Dementia Research & Education Centre, Hobart, Australia (2) Leeds Teaching Hospitals NHS Trust UK.

Word count: 3099
Number of Tables: 3
Number of Figures: 3
Number of Videos: 1

Corresponding Author:
Stefan Williams
Leeds Institute of Health Science, 10th Floor, Worsley Building
University of Leeds, Clarendon Way, Leeds, UK LS2 9LU
Phone: 0113 343 0839
umswi@leeds.ac.uk
ORCID: 0000-0002-4786-1356

Keywords:
Parkinson’s disease; Parkinsonism; Bradykinesia; Finger tapping; Computer Vision; Artificial Intelligence; DeepLabCut
ABSTRACT

Objective

The worldwide prevalence of Parkinson’s disease is increasing. There is urgent need for new tools to objectively measure the condition. Existing methods to record the cardinal motor feature of the condition, bradykinesia, using wearable sensors or smartphone apps have not reached large-scale, routine use. We evaluate new computer vision (artificial intelligence) technology, DeepLabCut, as a contactless method to quantify measures related to Parkinson’s bradykinesia from smartphone videos of finger tapping.

Methods

Standard smartphone video recordings of 133 hands performing finger tapping (39 idiopathic Parkinson’s patients and 30 controls) were tracked on a frame-by-frame basis with DeepLabCut. Objective computer measures of tapping speed, amplitude and rhythm were correlated with clinical ratings made by 22 movement disorder neurologists using the Modified Bradykinesia Rating Scale (MBRS) and Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS).

Results

DeepLabCut reliably tracked and measured finger tapping in standard smartphone video. Computer measures correlated well with clinical ratings of bradykinesia (Spearman coefficients): -0.74 speed, 0.66 amplitude, -0.65 rhythm for MBRS; -0.56 speed, 0.61 amplitude, -0.50 rhythm for MDS-UPDRS; 0.69 combined for MDS-UPDRS. All p < 0.001.

Conclusion

New computer vision software, DeepLabCut, can quantify three measures related to Parkinson’s bradykinesia from smartphone videos of finger tapping. Objective ‘contactless’ measures of standard clinical examinations were not previously possible with wearable sensors (accelerometers, gyroscopes, infrared markers). DeepLabCut requires only conventional video recording of clinical examination and is entirely ‘contactless’. This next generation technology holds potential for Parkinson’s and other neurological disorders with altered movements.
INTRODUCTION

The worldwide prevalence of Parkinson’s is increasing, with an estimated 10 million people already affected [1]. There is urgent need to find new tools to objectively measure the condition, to assist diagnosis, monitoring, and research metrics [2]. This is especially pressing in the era of precision medicine and with a global shortage of neurologists [3].

The cornerstone of Parkinson’s assessment is the clinician’s visual judgement of bradykinesia: one of the cardinal motor features of the condition, which is defined as slowness of movement and decrement in size or speed, or progressive hesitations/halts, as movements are continued [4]. There are several clinical examination methods to assess bradykinesia, but in a brief clinic room assessment, one of the most common is to observe the patient repeatedly tapping their index finger against thumb ‘as quickly and as big as possible’ [5]. However, clinicians’ visual interpretation is inherently subjective [6,7], with considerable inter-rater variability [5,9-11]. Previous reported methods to measure bradykinesia on finger tapping have required wearable equipment such as gyroscopes [9], electromagnetic sensors [12; 13], and infrared camera markers [14], or patient interaction with an app-based task [15], and this has hindered the pathway to large-scale, routine clinical use.

Recent developments in computer interpretation of video (computer vision), clear the pathway to ‘contactless’ methods of quantifying clinical assessments [16], with potential widespread global application because of the availability of inexpensive smartphone cameras [17]. DeepLabCut is recently published open source computer vision software, without a requirement to write new computer code, that has been demonstrated to track and measure mice and fruit fly body parts [18,19]. We apply it to clinical neurology for the first time and assess whether the technology can be used to track and measure bradykinesia in standard smartphone video of conventional finger tapping examination.

METHODS
Ethical review

The study was approved by the North of Scotland Research Ethics Committee, United Kingdom Health Research Authority (IRAS project ID 256116).

Participants and video recording

Video recordings of 138 hands (including left and right) were obtained from 39 patients with idiopathic Parkinson’s disease and 30 healthy controls, who gave written consent. One video was rejected because the hand moved outside the video frame, making 137 videos: 77 Parkinson’s hands and 60 control hands.

All Parkinson’s patients had previously been diagnosed by a movement disorder specialist neurologist at Leeds Teaching Hospitals NHS Trust United Kingdom, according to Movement Disorder Society clinical diagnostic criteria [4], and were subjectively and objectively in the ‘on’ state at the time of video recording (no medication was withheld prior to recording). We did not exclude patients with postural hand tremor or dystonia. Controls were recruited from the companions of patients, or hospital/university staff, and had no history of Parkinson’s disease or other neurological diagnosis.

Participants rested their elbow on a chair arm with the forearm lifted to a 45 degree angle. The hand was free to move as per the protocol of the Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) Item 3.4 Finger Tapping [5]. Only the hand/forearm was within the video frame. The distance from camera to hand was approximately 1m, but not tightly defined. Digits 1 and 2 were closest to the camera. No specific instructions were given for the position of digits 3 to 5, and participants were free to position these digits as they preferred, although the researcher gave a brief demonstration in which digits 3 to 5 were extended.
A smartphone, placed on a tripod, was used to record standard video (60 frames per second, 1920x1080 pixel resolution), with only ambient lighting. Participants were instructed to tap their index finger and thumb together “as quickly and as big as possible” for at least 10 seconds [5]. Each video was edited to an 11 second clip: 1 second prior to tapping onset and 10 seconds of finger tapping.

Clinical rating

22 consultant neurologists specialising in movement disorders (United Kingdom) were asked to rate 30 tapping videos each, selected at random from the set of 137 videos. The median number of raters per video was 5 (range 1 to 12, interquartile range 3 to 7). The raters were blinded to patient / healthy control status and each other’s scores. Each video was rated according to both the MDS-UPDRS Item 3.4 Finger Tapping [5], and the Modified Bradykinesia Rating Scale (MBRS) [8,9]. The MDS-UPDRS requires the rater to amalgamate judgments of finger tapping speed, amplitude, and rhythm into a single composite score (Table 1). In contrast, the MBRS is comprised of three separate scores for speed, amplitude, and rhythm (Table 2). The modal clinician rating for each video was used for correlation with computer measures (see below). For each possible pair of raters, Cohen Kappa was calculated based on all videos they had both seen. The resulting values had mean 0.28 (fair agreement) and standard deviation 0.28.

DeepLabCut video tracking

DeepLabCut tracks the geometrical configuration of multiple body parts in video, without a requirement to wear markers [18,20]. It uses transfer learning, with feature detectors based on deep neural networks that have been pretrained on ImageNet, a massive object recognition dataset [21]. This means that DeepLabCut is able to accurately recognise and track body parts with minimal training data [18].

The set of 137 finger tapping videos was processed by DeepLabCut. We localised (labelled) six distinct points on the tapping hand in 20 frames selected by k-means clustering from each 660 frame video:
thumb tip; index finger tip; thumb metacarpophalangeal (MCP) joint; index finger MCP joint; middle
finger tip; dorsal wrist/proximal dorsal hand. The deep neural network architecture of DeepLabCut was
then trained using these points to predict their localisation in the remaining 97% of (unlabelled) video
frames (1030000 training iterations, ResNet 50). This created video pixel coordinates for finger tip and
thumb tip throughout each video.

The accuracy of DeepLabCut to track the hand localisation points (including finger tip and thumb tip)
was assessed using the ‘evaluate network’ function within DeepLabCut. This function computes the
Euclidean error between the manual labels and the ones predicted by DeepLabCut averaged over the
hand locations and test images (mean absolute error, proportional to the average root mean square error)
[20].

Signal processing

Video pixel coordinates for the labels produced by DeepLabCut were used to calculate the pixel distance
between index finger tip and thumb tip. A Savitzky-Golay filter was applied to the resulting time series,
which removed large, sudden transient label ‘jumps’ caused by DeepLabCut mislabelling (i.e. large-
distance, physiologically impossible label movement across one pair of video frames). The pixel
number distances were standardised across all videos, by using the maximum opening distance between
finger and thumb tip and normalising this to a value of 1 (all values were divided by the maximal value
in the corresponding time series).

Three features of the resulting finger tip to thumb tip distance time series were calculated to reflect the
clinical features of finger tapping speed, amplitude and rhythm (Figure 1). A measure of speed was
calculated as the mean rate of change of the normalised distance between finger and thumb tip over
time. A measure of amplitude variability was calculated by dividing each time series into one-second
windows, with maximal overlap, and then calculating the coefficient of variation of the mean difference
between the maximum and minimum amplitudes in each window. A measure of rhythm regularity was
calculated by undertaking Fast Fourier Transform to find the distribution of frequencies within each
finger tap time series, and then measuring the power of the dominant frequency peak added to the power
of the frequencies 0.2 Hz either side of it. A more regular tapping rhythm concentrates power in a
narrow frequency band, increasing the power of the dominant frequency peak (and its immediate
neighbours), whereas a more irregular tapping rhythm leads to a more widely spread distribution of
frequency bands, reducing the power of the dominant frequency peak (and its immediate neighbours).

Spearman correlation coefficients were calculated for the each of the three DeepLabCut measures vs
the modal MBRS and MDS-UPDRS clinician ratings. In addition, we calculated the Spearman
correlation coefficient of the three computer scores combined (the normalised arithmetic mean of the
computer speed score, amplitude variability score and rhythm regularity score) with the MDS-UPDRS
rating.

RESULTS

The participant and hand video characteristics are given in Table 3.

Tracking accuracy

DeepLabCut reliably tracked the finger tapping movements across 133 videoed clinical examinations.
Figure 2 shows example frames from a video labelled by DeepLabCut. The mean absolute error of
DeepLabCut labelling was 8.39 pixels, i.e. the average distance between manual (human) labels and
those predicted by DeepLabCut was 8.39 pixels within a 1920x1080 pixel video frame. Video 1 shows
examples of four videos labelled by DeepLabCut, including particularly challenging cases with tremor
and dystonia also present.

Correlation with clinical bradykinesia ratings
The quantitative scores derived from DeepLabCut measurements correlated well with the group neurologists’ modal ratings of bradykinesia; Figure 3. The computer vision measures of tapping speed, amplitude variability and rhythm regularity had good correlations with the respective MBRS clinical ratings for speed (-0.70, p < 0.001), amplitude (0.65, p < 0.001) and rhythm (-0.61, p < 0.001). There were also good correlations between the computer measures of tapping speed (-0.66, p < 0.001), amplitude variability (0.56, p < 0.001) and rhythm regularity (-0.50, p < 0.001) and the MDS-UPDRS clinical rating. The MDS-UPDRS is a composite clinical rating, combining the separate components of bradykinesia, and the three computer measures combined also correlated with MDS-UPDRS (-0.69, p < 0.001).

DISCUSSION

We have shown that new computer vision software, DeepLabCut, can track finger tapping in a standard smartphone video recording of a conventional clinical examination, and quantify three measures related to Parkinson’s bradykinesia: speed, amplitude variation and rhythm regularity. We found good correlations between the objective computer measures of finger tapping and clinical ratings of bradykinesia by 22 movement disorder neurologists using two gold standard clinical scales (MBRS and MDS-UPDRS). Thus, smartphone video measures using DeepLabCut exhibit convergent validity with conventional clinician ratings.

These results are new and clinically significant. This is the first study to apply new open source DeepLabCut technology, developed for animal behaviour studies, to the field of human movement disorders, or indeed any aspect of clinical neurology. We demonstrate new objective measures related to bradykinesia by bringing together two ubiquitous ingredients: the smartphone video camera and visual artificial intelligence software. With the worldwide ownership of smartphones so high (e.g. 82% in the UK, 77% in the US) [17], there are no hardware barriers to further use of this approach.
Neurologists recognize that quantification of bradykinesia will improve Parkinson’s management through ‘precision medicine’ principles: it has the potential to aid early accurate diagnosis and remote monitoring, and provide fine grain regular measurement of clinical trial outcomes. Previous methods to quantify bradykinesia have focussed on wearable sensors measuring finger tapping [9,12-14,23] but it has been challenging to incorporate these into widespread use because equipment needs to be physically attached to the patient. Furthermore, the necessity for sensors to stay fixed on a specific body part has often required the exclusion of tremulous and dyskinetic patients [8,9,12,14,23]. In contrast, the DeepLabCut test presented here is a contactless method that simply requires a short 10 second video recording of the standard clinical examination. This means it can potentially be used in any setting and any patient.

The ‘real life’ versatility of DeepLabCut in ordinary clinical settings was evaluated in this study because no special efforts were made to optimize participant positioning or filming conditions. The tracking and measures were robust despite variations in smartphone camera distance, positioning of the hand, skin color, flexion / extension postures of digits 3 to 5, ambient lighting conditions, and inclusion of a small number of patients with tremor or dystonia during tapping. Video 1 illustrates this, with accurate labelling of finger and thumb tip despite superimposed tremor or dystonia. This is particularly encouraging for the relevance of DeepLabCut aiding assessment of Parkinson’s in conventional clinical settings.

Two further strengths of our study are the collection of clinical ratings from a large group of movement disorder neurologists, and the use of two rating scales: the MBRS and MDS-UPDRS, such that the clinical ratings were representative and reliable. Clinical rating scales are vulnerable to inter-rater variability but using modal values derived from a large number of blinded clinical raters, and two different gold-standard rating scales, provided a more robust ‘ground truth’ of clinical bradykinesia to evaluate the new test against. As such we were able to demonstrate that clinician and DeepLabCut measures correlate for both individual components of bradykinesia (MBRS) and an overall combination.
of those components (MDS-UPRDS). An interesting future application with a larger participant group would be to examine the relative contribution of the three video measures to individual grades of UPDRS bradykinesia.

Another advantage of the approach described here is that it is not limited by the physical constraints of wearable or screen/keyboard tapping methods, and could be applied to detect and measure other motor features of Parkinson’s in a contactless manner such as tremor, gait, or posture. Computer vision is general has potential for application to these signs [24,25].

Only three previous studies have objectively measured finger tapping using contactless, standard video analysis. One method extracted finger tapping information from 13 Parkinson's patients with advanced disease [26], but it required inclusion of the patient's face in the video, limiting the practical utility. The second report, from our own group, described a video method to measure movement using an optical flow field [27] that interprets movements of the whole hand. The work here improves on this by explicitly extracting the salient thumb and finger tip points, and validating in a larger cohort. A third study of 60 Parkinson’s patients (no controls) found strong correlation between MDS-UPRDS finger tap rating and video measures of tap interval (frequency) \((r=0.91)\), frequency variation \((r=0.82)\) and amplitude \((r=-0.94)\) but not amplitude variation \((r=0.39)\) [28]. The ‘ground truth’ of the clinical comparison was less robust, based on just two clinical raters, and there were no MBRS ratings; furthermore, some images were blurred in the 24 frame per second videos. A major advantage of our method in comparison is that DeepLabCut is open source software, that can be downloaded and used without the need to write computer code, making it available now to neurologists with only limited computing knowledge.

A well-studied, non-camera method to measure finger tap bradykinesia is the ‘Bradykinesia-Akinesia Incoordination (BRAIN) Test’, which involves participants tapping a standard computer keyboard for 30 seconds [29]. This shares advantages with our method in that it involves no special equipment, and could be employed without clinician-patient contact. However, it requires patient motivation to engage
with a keyboard tapping task, rather than a simple recording of existing standard clinical examination (and similar requirements for patient motivation apply to patient smartphone apps [15]). Furthermore, ‘BRAIN’ correlations with MDS-UPDRS finger tapping ratings of 0.44, 0.20, 0.33, 0.03 [29], are weaker than those we report (0.66, 0.56, 0.50, 0.69). One reason for this may be that a computer keyboard cannot record any measure of tap amplitude, and the speed measure is limited to tap frequency, so that some core aspects of bradykinesia cannot be captured by ‘BRAIN’.

The present study has some limitations. Like human vision, a simple camera lacks an absolute measure of distance, and our method cannot capture 3D movement. Rotation of the thumb and finger into a horizontal plane would falsely alter amplitude measurement. However, in practice such movement is rare during tapping, and our normalized measure of finger-thumb distance based on 2D maximum opening distance appeared to capture amplitude, speed and rhythm abnormalities well, given the good correlations with clinical ratings.

We have measured amplitude variation, but not amplitude decrement (the sequence effect). Previous approaches have measured decrement as a straight line linear regression [23], but decrement would also proportionately affect a measure of amplitude variation, such that the variations in decrement should be captured within our measure. Furthermore, there are several other previous reports of good correlation between non-decrement amplitude measures using sensors and clinician ratings of finger tap bradykinesia [9].

Relatively few videos were rated as grade 3 or 4 bradykinesia, and it is possible that a greater number of higher grade videos might change the strength of correlations. Even with accurate tracking, all forms of rating or measuring bradykinesia (including clinician) may potentially be confounded by comorbidities such as joint deformities, pain, dystonia etc. However, with a larger dataset, machine learning techniques are well-suited to separate such additional contributions from underlying bradykinesia. There was a degree of error in DeepLabCut labelling of finger and thumb tips, but this was minimal (mean of 8.39 pixels within a 1920x1080 pixel video frame) and unlikely to affect
individual measures. We mitigated against this by applying a filter prior to feature extraction. Full
devision of digits 3 to 5 might improve labelling accuracy, but would also limit the amplitude of finger
tapping [23], potentially masking more subtle deficits. The labelling error provided by DeepLabCut’s
‘evaluate network’ function only gives an accuracy figure for the entire set of videos but not for
individual videos, meaning that we have not reported accuracy in subsets of videos, such as those with
visible tremor. This study was exploratory, aimed at proof of concept that bradykinesia related measures
could be derived from smartphone video. The protocol was thus simple, and repeat measures such as
test-retest in controls, or ‘off’ and ‘on’ medication states, were not included, meaning that reliability
and sensitivity to change have not been characterised. Future work will assess this.

Finally, we have demonstrated measurements that show convergent validity with clinical ratings, rather
than a test to discriminate patients from controls. In our view, it makes little sense to pursue automated
diagnostic classification using finger tapping video alone because clinical Parkinson’s diagnosis is by
definition a broader assessment than isolated finger tapping performance [4].

This study has evaluated the clinical application of DeepLabCut in a medical condition, specifically a
neurological disorder. The methodology described here provides a new objective measure of
Parkinson’s bradykinesia in conventional clinical examination. DeepLabCut provides a contactless
means for tracking and measuring bradykinesia in Parkinson’s, without any special equipment other
than the ubiquitous smartphone. This research therefore provides a new tool to quantify bradykinesia.
It could potentially be used to support diagnosis and monitoring of Parkinson’s, both clinically and as
an outcome measure in research studies. Beyond Parkinson’s and other movement disorders, this same
technology could be applied to many other neurological conditions that require measurement of
movements or function; for example multiple sclerosis, neuropathies, epilepsy and stroke.

REFERENCES


**Conflict of Interest Statement**

The authors declare that they have no conflict of interest.

**FIGURE LEGENDS**
**Figure 1.** Illustration of the three parameters derived from the DeepLabCut finger tip and thumb tip coordinates to measure finger tapping speed, amplitude and rhythm.

**Figure 2.** Example video frames taken from smartphone video labelled by DeepLabCut, showing the six localization (labelling) points: thumb tip (dark blue), thumb metacarpophalangeal (MCP) joint (light blue), index finger tip (cyan), index finger MCP joint (yellow), middle finger tip (orange), dorsal wrist (red).

**Figure 3.** Computer measures of finger tapping speed, amplitude variation and rhythm, derived from DeepLabCut smartphone video tracking, correlate with clinician mode ratings for the same videos. Pearson correlation coefficients shown. MBRS: Modified Bradykinesia Rating Scale. MDS-UPDRS: Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale.

### Tables

**Table 1.** The Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) Item 3.4 (Finger Tapping).

<table>
<thead>
<tr>
<th>MDS-UPDRS Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: Normal</td>
<td>No problems</td>
</tr>
<tr>
<td>1: Slight</td>
<td>Any of the following: a) the regular rhythm is broken with one or two interruptions or hesitations of the tapping movement; b) slight slowing; c) the amplitude decrements near the end of the 10 taps.</td>
</tr>
<tr>
<td>2: Mild</td>
<td>Any of the following: a) 3 to 5 interruptions during tapping; b) mild slowing; c) the amplitude decrements midway in the 10-tap sequence.</td>
</tr>
<tr>
<td>3: Moderate</td>
<td>Any of the following: a) more than 5 interruptions during tapping or at least one longer arrest (freeze) in ongoing movement; b) moderate slowing; c) the amplitude decrements starting after the 1st tap.</td>
</tr>
<tr>
<td>4: Severe</td>
<td>Cannot or can only barely perform the task because of slowing, interruptions, or decrements.</td>
</tr>
</tbody>
</table>
Each hand is tested separately. The patient is instructed to tap the index finger on the thumb “as quickly and as big as possible”. The first 10 taps are evaluated by the clinician [5].

Table 2. The Modified Bradykinesia Rating Scale (MBRS).

<table>
<thead>
<tr>
<th>Score</th>
<th>Speed</th>
<th>Amplitude</th>
<th>Rhythm</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal</td>
<td>Normal</td>
<td>Regular, no arrests or pauses in ongoing movement</td>
</tr>
<tr>
<td>1</td>
<td>Mild Slowing</td>
<td>Mild reduction in amplitude in later performance, most movements close to normal</td>
<td>Mild impairment, up to two brief arrests in the 10 seconds, none lasting &gt; 1 second</td>
</tr>
<tr>
<td>2</td>
<td>Moderate Slowing</td>
<td>Moderate, reduction in amplitude visible early in performance but continues to maintain 50% amplitude through most of the tasks</td>
<td>Moderate, 3 to 4 arrests in 10 seconds; OR 1 or 2 lasting &gt; 1 second</td>
</tr>
<tr>
<td>3</td>
<td>Severe slowing</td>
<td>Severe, less than 50% amplitude through most of the task</td>
<td>Severe, 5 or more arrests/10 seconds; OR more than 2 lasting &gt; 1 second</td>
</tr>
<tr>
<td>4</td>
<td>Can barely perform the task</td>
<td>Can barely perform the task</td>
<td>Can barely perform the task</td>
</tr>
</tbody>
</table>

Each hand is tested separately. The patient is instructed to tap the index finger on the thumb “as quickly and as big as possible”. The first 10 seconds of tapping are evaluated by the clinician to give three MBRS ratings (one for each of speed, amplitude, rhythm) [6,7].

Table 3. Study participant (hand video) characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Patients</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Std. Dev.) yrs</td>
<td>68 (9.6)</td>
<td>59 (19.4)</td>
</tr>
<tr>
<td>Male/Female</td>
<td>47/26</td>
<td>22/38</td>
</tr>
<tr>
<td>Median years since diagnosis</td>
<td>4</td>
<td>n/a</td>
</tr>
<tr>
<td>Median H&amp;Y [IQR]</td>
<td>2 [1,3]</td>
<td>n/a</td>
</tr>
<tr>
<td>H&amp;Y = 1</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 1.5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 2</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 2.5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 3</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>H&amp;Y = 4</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
Data is split by Parkinson’s hands (n=73) and control hands (n=60). H&Y: modified Hoehn and Yahr scale [22]. MDS-UPDRS: Movement Disorder Society revision of the Unified Parkinson’s Disease Rating Scale, Item 3.4 (Finger Tapping). IQR: Interquartile Range.

Acknowledgements

We would like to thank the following UK movement disorder neurologists for providing clinical rating of videos: Dr Jane Alty, Dr Sundus Alusi, Dr Michael Bonello, Dr Stephen Butterworth, Dr Philip Buttery, Dr Camille Carroll, Dr Adam Cassidy, Dr Jeremy Cosgrove, Dr Richard Ellis, Dr Jonathan Evans, Dr Paul Goldsmith, Dr Donald Grosset, Dr Christopher Kobylecki, Dr Alistair Lansbury, Dr Peter Moore, Dr Rachel Newby, Dr Edward Newman, Dr Gillian Sare, Dr Monty Silverdale, Dr Naomi Warren, Dr Louise Wiblin, Dr Caroline Williams-Gray.

We would also like to thank Dr Parisa Patel and Dr Paschal O’Gorman for data collection.

Author contributions

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Role</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Institution and Affiliation</td>
<td>Authorship</td>
<td>Contributions</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Stefan Williams, MBBS, BSc</td>
<td>University of Leeds, UK</td>
<td>Author</td>
<td>Design and conceptualized study; data acquisition and interpretation; drafted the manuscript for intellectual content</td>
</tr>
<tr>
<td>Zhibin Zhao, BEng</td>
<td>(1) University of Manchester, UK; (2) Xi’an Jiatong University, China</td>
<td>Author</td>
<td>Analyzed the data (major role)</td>
</tr>
<tr>
<td>Awais Hafeez, MA, BSc</td>
<td>(1) University of Leeds, UK; (2) University of Engineering and Technology Lahore, Pakistan</td>
<td>Author</td>
<td>Design and conceptualized the study; major role in acquisition of data</td>
</tr>
<tr>
<td>David C Wong, PhD</td>
<td>University of Manchester, UK</td>
<td>Author</td>
<td>Design and conceptualized study; analyzed the data; data interpretation; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Samuel D Relton, PhD</td>
<td>University of Leeds, UK</td>
<td>Author</td>
<td>Design and conceptualized study; data interpretation; revised the manuscript for intellectual content</td>
</tr>
<tr>
<td>Hui Fang, PhD</td>
<td>Loughborough University, UK</td>
<td>Author</td>
<td>Design and conceptualized study (major role)</td>
</tr>
<tr>
<td>Jane E Alty, MD, MA, MBBChir</td>
<td>(1) University of Tasmania, Australia; (2) Leeds Teaching Hospitals NHS Trust, UK</td>
<td>Author</td>
<td>Design and conceptualized study; data interpretation; drafted the manuscript for intellectual content</td>
</tr>
</tbody>
</table>
**Figure 1**

Speed Measure: mean rate of change of finger-thumb distance

Amplitude Variability Measure: mean peak-to-peak value calculated, then coefficient of variation calculated

Finger-thumb distance

**Fast Fourier Transform**

Rhythm Regularity Measure: a more regular rhythm will create a narrower frequency distribution with a higher dominant peak
Figure 3

(a) \( r = 0.70, p \ll 0.001 \)

(b) \( r = 0.65, p \ll 0.001 \)

(c) \( r = -0.62, p \ll 0.001 \)

(d) \( r = -0.65, p \ll 0.001 \)

(e) \( r = 0.57, p \ll 0.001 \)

(f) \( r = -0.44, p \ll 0.001 \)
Click here to access/download
e-component
Video_1.MOV