Artificial intelligence analysis of videos to augment clinical assessment: an overview

David C. Wong[*], Stefan Williams

Observation is a fundamental part of the practice of clinical medicine. Observation of movement is particularly important for the neurologist. Conditions such as Parkinson's disease, multiple sclerosis, stroke, epilepsy, and many others affect a person's movement in characteristic ways. In some conditions, changes in the patient's voice can be included in this – changes in sound caused by changes in the movements of speech. The clinician's detection of a characteristic abnormality, and their judgment of its severity, plays a central role in both diagnosis and the assessment of prognosis or response to treatment. However, that practice depends upon a limited resource of experienced experts. In addition, these experts are limited by human visual judgment, which cannot reliably or precisely detect and measure small or subtle changes in movement (Williams et al., 2023).

Researchers have previously attempted to address these limitations by developing a range of sensors for measuring the changes in movement that occur in neurological conditions. These include the accelerometer, gyroscope, and full motion capture. Each allows measurement of motor function and has been shown, in multiple studies, to provide an accurate assessment of motor signs (Memar et al., 2018). However, all are limited in that they require special equipment. Such equipment would be expensive at the scale required, and so is not part of routine care, nor available in many global health contexts. Additionally, specific sensors are limited to measuring limb motor function, and cannot easily detect more subtle clinical indicators such as changes in facial expression and voice.

A promising solution to these problems is to assess patient movement using visual and audio information from smartphone video using artificial intelligence (AI) methods - a mixture of computer vision and audio signal processing methods that we refer to here as Video AI. This has the potential to extract clinically useful information that can augment and improve diagnosis and assessment. Unlike other sensors, smartphones are commonly used by the general public across the globe, in both developed and low- and middle-income countries. Videos are non-contact, and may therefore be suitable in situations with higher risk of infection. Videos can also be recorded remotely, without the presence of a trained clinician or technician, and may be useful within mobile health applications. This nascent technology has the potential to fundamentally shift the practice of clinical neurology.

Examples of video-based AI for augmenting clinical assessment and monitoring of neurological conditions:
Here, we provide three examples of how such technology has been proposed for monitoring and measuring disease symptoms and progression and explain how it might be used in different parts of the clinical pathway for common neurological conditions. The examples were chosen to illustrate some of the principles and potential of video-based artificial intelligence methods, in common neurological conditions.

Recognition and quantification of hand tremor:
Hand tremor is a common sign of neurological conditions such as Essential Tremor, Parkinson's disease, or Functional Neurological Disorder. To assess the cause and severity of a tremor, a clinician observes its amplitude and frequency for different hand positions and actions. Video can be used to objectively measure both of these features of tremor: the amplitude and the frequency. Using neural networks that have been trained to identify and locate objects, we can track hands and individual digits within a video over time (**Figure 1**). Such networks are typically trained on very large data sets, and the resulting models have been released publicly. Examples include Openpose, DeepLabCut, and Mediapipe. These publicly-available models are sometimes retrained on a local dataset, a process known as transfer learning. Such methods have been shown to be robust to lighting conditions and physical environment, such that they could be used in real-life situations (Huang et al., 2022). By tracking the motion of landmarks on the hand over time, we can measure hand tremor frequency by quantifying periodic movements using Fourier analysis. For video recordings of patients with tremors, Williams et al. (2021) show that video measurement of tremors has a near-perfect agreement with accelerometer measurements.

Hand motion video data are insufficient, by itself, to measure tremor amplitude as it is impossible to distinguish between hands that are large or close to the video camera. However, by assessing the distance between the hand and camera, for instance, by using the smartphone's in-built depth camera, tremor amplitude can also be measured in distance units (Bungay et al., 2023). Bungay's pilot data showed agreement within 1.2 cm in most cases, in comparison to a ruler that measured to the nearest 0.5 cm. This approach is limited, as tremor is assumed to be a fixed distance from the camera (i.e., in the image plane). Full 3D hand point tracking using smartphones ought to be possible, given that the most recent smartphone depth cameras can record at a high frame rate (60 fps) and resolution (1080 p). This would allow more faithful measurement of Parkinsonian tremors that can include rotational, as well as lateral, motion.

Automatic classification of finger-tapping bradykinesia:
Bradykinesia is the cardinal motor sign of Parkinson's disease, present in every person with the condition. It is defined as slowness of movement and decrement in amplitude or speed (or progressive hesitations/halts) as movements are continued. Bradykinesia is tested by observation of repetitive movement, such as finger-thumb tapping, in which the patient is asked to tap finger and thumb together as quickly and as big as possible. The neurologist must visually judge whether bradykinesia is present, but also its severity. In research settings, including trial outcomes, finger-tapping bradykinesia severity is graded on a five-point rating scale using the Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS). Based on observation of 10 consecutive taps, the score ranges from 0 (no bradykinesia) to 4 (severe bradykinesia), depending on the degree of slowing, number of hesitations/halts, or how early a decrement in tapping amplitude or speed occurs.

Just as a neurologist will observe the relative motion of the thumb and forefinger, one can use the same finger tracking method used in tremor measurement to track the thumb and index finger tips during the tapping test. By plotting the distance between thumb and finger, one can derive metrics of the size, speed, rhythm, and decrement in rhythm (Williams et al., 2020; Zhao

et al., 2020) corresponding to the components of bradykinesia. The correlation between the mean of these metrics and the MDS-UPDRS score was 0.69, indicating moderate convergent validity of the computer vision method.

Morinan et al. (2023) have recently extended this concept by analyzing multiple clinical tests from multiple limbs, per participant, over a much larger multi-site data set. They showed that the combination of information from these tests can be used to estimate a score of whole-body bradykinesia. The score was calculated as the sum of individual limb tests, each scored using 0–4 MDS-UPDRS categories, for a total scale between 0 and 40 (Morinan et al., 2023). In comparison between clinician-assessed and computerized scores, they report an interclass correlation of 0.74, and that in 84% of assessments, disagreement between the clinician and the model was within ± 7 of the true score, which is considered to be the threshold for a large clinically important difference.

Although we have discussed examples involving hand tracking, this general approach involving markless tracking has been explored for other body parts. For instance, whole-body tracking has been explored to assess Parkinsonian gait and face markers could be used to determine facial muscle control.

Parkinson's disease progression from voice recordings:
Video data contains both visual and audio signals. The audio signal may capture clinical information that cannot be adequately captured in just the visual signal. For instance, Parkinson's disease affects speech in multiple ways, including articulation, phonation, and speech fluency. These abnormalities are collectively referred to as hypokinetic dysarthria. One approach has used sound recordings of people sustaining the /ah:/ vowel sound (Tsanas et al., 2021). By extracting key characteristics of the sound, such as change in pitch, Tsanas et al. (2021) developed a model that could predict overall Parkinson's severity according to the Unified Parkinson's Disease Rating Scale (UPDRS) score to within 3.5 points, in comparison to clinical gold standard assessment.

The approaches used in each of the three described examples first extract pertinent information from the raw video, guided by clinical insight. For instance, in the case of finger-tapping bradykinesia, the rhythm and speed of the thumb and forefinger are first extracted. However, we have seen, in other clinical situations, how modern deep-learning approaches are able to derive clinical insights directly from the raw data. Such methods have not been feasible until recently, as they require large quantities of video data for model training, but we highlight this as a potential approach for further improvements (**Figure 1**).

Potential clinical usage:
We envision these technologies having multiple possible uses throughout all stages of clinical care.

Prior to diagnosis, video AI methods may be useful in assisting screening and triage for onward referral at the primary care level. Most symptoms, such as a problem with gait, or a tremor, have multiple potential causes. Video AI could help primary care physicians more accurately judge the urgency or necessity of onward referral to a neurologist. For example, it might be possible to characterize a tremor as low risk for Parkinson's disease based on its visual features and a combined analysis of finger tapping and other video. This would be useful in primary care, where clinicians are less likely to have regular experience with specific neurological tests.

If quantitative measures of visible signs are sufficiently discriminatory between conditions, it is also possible they could move beyond assistance with triage and more directly aid with diagnosis in secondary care. Whilst current attempts at differential diagnosis from user-entered symptoms have so far been unreliable (Fraser et al., 2018), the use of detailed visual and audio information,

Perspective

David C. Wong[*], Stefan Williams
University of Leeds; Stefan Williams, Leeds
Teaching Hospitals Trust, Leeds, UK
***Correspondence to:** David C. Wong, DPhil,
d.c.wong@leeds.ac.uk.
https://orcid.org/0000-0001-8117-9193
(David C. Wong)

Open access statement: *This is an open access journal, and articles are distributed under the terms of the Creative Commons AttributionNonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.*

Open peer reviewer: *Peter B Marschik, University Medical Center Göttingen, Germany; Tamás Karácsony, University of Porto, Portugal.*

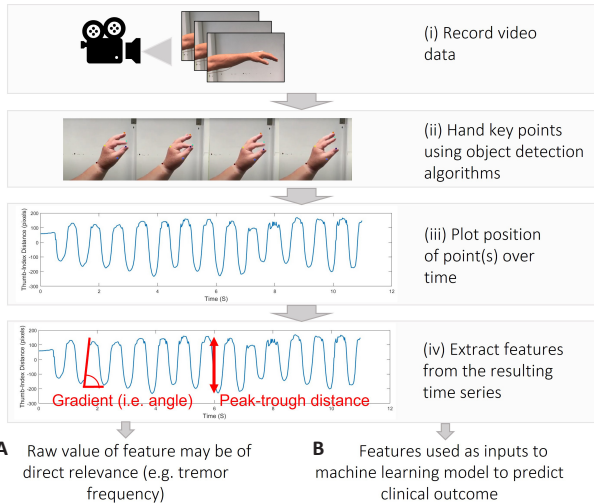**Additional file:** *Open peer review reports 1 and 2.*

Figure 1 │ An example pipeline showing how coordinates of pre-determined clinically relevant points can be tracked and extracted in two dimensions from video, or in three dimensions if a camera with a depth sensor is used.
The time series of points is then used to either (A) directly provide objective measurements, (B) as inputs to a machine learning model to predict some clinical outcome. (C) Alternatively, raw video can be used to train deep learning models to predict a clinical outcome. The figure was drawn using Microsoft PowerPoint, and graphs were plotted in Python using Pyplot.

in combination with patient-reported information, better reflects the information available to a clinician and has the potential to aid more accurate and reliable diagnosis.

Once a neurological disorder has been diagnosed, it is important to monitor progression or change over time. This enables prompt recognition of the need for treatment initiation or escalation, the effectiveness of a treatment change, and an idea of prognosis based on change over time. In current clinical practice, this type of post-diagnostic assessment is limited to intermittent face-to-face appointments, and relies on both subjective patient-reported measures and clinical observations. Instead, it would be possible to record video metrics to augment existing patient symptom diary apps.

Video AI may also be used to augment video consultations, helping to reduce the burden of unnecessary travel in patients that often suffer a degree of mobility impairment. We note that the uptake of video consultations has thus far been poor, as they are perceived to offer little practical value over telephone consultations (Greenhalgh et al., 2022). However, the approaches described necessitate the use of both audio and visual information only available via video.

Many neurological research trials involve neurological examination signs as part of outcome measures. The researchers apply clinical rating scales, in which they must make a visual judgment, for example, the amplitude of tremor or the severity of bradykinesia according to five-point rating scales (TETRAS, MDS-UPDRS finger-tapping score, etc). Video AI offers the potential to automate this outcome measurement, potentially providing more sensitive and reliable means to detect early or subtle effects that direct treatment development.

Implementation challenges: Although video-based methods hold high promise, embedding these into clinical practice is non-trivial. All of the use cases described here are related to direct patient care, and any resulting system would be deemed to be Software as a Medical Device in the US and UK. Any system would need to show evidence that it is sufficiently accurate for the task at hand. The methods described above have, for the most part, been developed in relatively small cohorts, and there remain uncertainties about their use in real-world conditions, or diverse populations, for instance, with different skin tones.

We note that collecting appropriate video data in

trials is itself a challenging task, due to both the size of the data and the difficulty in anonymizing it. Indeed, approaches that use whole-body or face videos especially need to consider patient acceptability, in addition to any legal requirements, of recording and storing readily identifiable information.

In addition, such technology can be considered a complex intervention involving multiple stakeholders and existing clinical pathways. Such interventions are prone to failure due to a lack of consideration of the wider clinical landscape into which it is embedded. For instance, the use of this technology within a patient self-monitoring mobile health application would need to consider whether the target population are suitably trained and supported.

Recent guidance from the DECIDE-AI collaborative (Vasey et al., 2022) highlights the importance of evaluations that include analysis of existing processes and infrastructure as well as the human-computer interaction between clinician and technology. Any patient-facing use cases, such as remote monitoring, must also consider the additional patient-clinician and patient-technology interactions. For instance, this may include providing technical support for patients, a role that may not exist under existing clinical pathways. DECIDE-AI's accompanying checklist provides guidance on the multiple facets that ought to be considered, evaluated, and reported in the initial stages of AI technology rollout.

Conclusion: We have highlighted several ways in which smartphones, in combination with AI analysis, have been used to measure and analyze the signs of neurological disease. Although there remain challenges in integrating these methods into routine clinical care, the ubiquity of the smartphone addresses one significant hurdle towards large-scale adoption. In the near term, they have the potential to objectively measure symptom characteristics so that subtle changes can be detected or monitored over time. Such objective measurements could augment clinical judgment in a way that produces more consistent clinical decision-making, potentially leading to better long-term management and outcomes for patients.

References

Bungay J, Emokpae O, Relton S, Alty J, Williams S, Fang H, Wong DC (2023) Contactless hand tremor amplitude measurement using smartphones: development and pilot evaluation. Conf Proc IEEE Eng Med Biol Soc doi:10.48550/arXiv.2304.14937.

Fraser H, Coiera E, Wong D (2018) Safety of patient-facing digital symptom checkers. Lancet 392:2263-2264.

Greenhalgh T, Ladds E, Hughes G, Moore L, Wherton J, Shaw SE, Papoutsi C, Wieringa S, Rosen R, Rushforth A, Rybczynska-Bunt S (2022) Why do GPs rarely do video consultations? qualitative study in UK general practice. Br J Gen Pract 72:e351-360.

Huang G, Tran SN, Bai Q, Alty J (2023) Real-time automated detection of older adults' hand gestures in home and clinical settings. Neural Comput Appl 25:8143-8156.

Memar S, Delrobaei M, Pieterman M, McIsaac K, Jog M (2018) Quantification of whole-body bradykinesia in Parkinson's disease participants using multiple inertial sensors. J Neurol Sci 387:157-165.

Morinan G, Dushin Y, Sarapata G, Rupprechter S, Peng Y, Girges C, Salazar M, Milabo C, Sibley K, Foltynie T, Cociasu I, Ricciardi L, Baig F, Morgante F, Leyland LA, Weil RS, Gilron R, O'Keeffe J (2023) Computer vision quantification of whole-body Parkinsonian bradykinesia using a large multi-site population. NPJ Parkinsons Dis 9:10.

Tsanas A, Little M, Ramig L (2021) Remote assessment of Parkinson's disease symptom severity using the simulated cellular mobile telephone network. IEEE Access 9:11024-11036.

Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, Denniston AK, Faes L, Geerts B, Ibrahim M, Liu X, Mateen BA, Mathur P, McCradden MD, Morgan L, Ordish J, Rogers C, Saria S, Ting DSW, Watkinson P, et al. (2022) Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ 377:e070904.

Williams S, Zhao Z, Hafeez A, Wong DC, Relton SD, Fang H, Alty JE (2020) The discerning eye of computer vision: Can it measure Parkinson's finger tap bradykinesia? J Neurol Sci 416:117003.

Williams S, Fang H, Relton SD, Wong DC, Alam T, Alty JE (2021) Accuracy of smartphone video for contactless measurement of hand tremor frequency. Mov Disord Clin Pract 8:69-75.

Williams S, Wong D, Alty JE, Relton S (2023) Parkinsonian hand or clinician's eye? Finger tap bradykinesia interrater reliability for 21 movement disorder experts. J Parkinsons Dis doi:10.3233/JPD-223256.

Zhao Z, Fang H, Williams S, Relton SD, Alty J, Casson AJ, Wong DC (2020) Time series clustering to examine presence of decrement in Parkinson's finger-tapping bradykinesia. Annu Int Conf IEEE Eng Med Biol Soc 2020:780-783.

P-Reviewers: Marschik PB, Karácsony T; C-Editors: Zhao M, Sun Y, Qiu Y; T-Editor: Jia Y