Deep Learning Parkinson's from Smartphone Data

C. Stamate*, G.D. Magoulas*, S. Kueppers†*, E. Nomikou‡,
I. Daskalopoulos*, M.U. Luchini†, T. Moussouri‡, and G. Roussos*
*Birkbeck College, University of London
†Benchmark Performance Ltd
‡Audience Focus Ltd

***Abstract*—The cloudUPDRS app is a Class I medical device, namely an active transient non-invasive instrument, certified by the Medicines and Healthcare products Regulatory Agency in the UK for the clinical assessment of the motor symptoms of Parkinson's Disease. The app follows closely the Unified Parkinson's Disease Rating Scale which is the most commonly used protocol in the clinical study of PD; can be used by patients and their carers at home or in the community; and, requires the user to perform a sequence of iterated movements which are recorded by the phone sensors. This paper discusses how the cloudUPDRS system addresses two key challenges towards meeting essential consistency and efficiency requirements, namely: (i) How to ensure high-quality data collection especially considering the unsupervised nature of the test, in particular, how to achieve firm user adherence to the prescribed movements; and (ii) How to reduce test duration from approximately 25 minutes typically required by an experienced patient, to below 4 minutes, a threshold identified as critical to obtain significant improvements in clinical compliance. To address the former, we combine a bespoke design of the user experience tailored so as to constrain context, with a deep learning approach used to identify failures to follow the movement protocol while at the same time limiting false positives to avoid unnecessary repetition. We address the latter by developing a machine learning approach to personalise assessments by selecting those elements of the UPDRS protocol that most closely match individual symptom profiles and thus offer the highest inferential power hence closely estimating the patent's overall UPRDS score.**

I. Introduction

Parkinson's Disease (PD) is a degenerative neurological condition associated with a wide spectrum of symptoms including tremor, slowness of movement and freezing, swallowing difficulty, sleep-related difficulties and psychosis [17]. Care for patients with PD involves the management of both motor and non-motor symptoms as well as palliative care. Since there is no cure, symptom management is a life-long process that affects not only the patients but also their families and carers. Clinical care pathways include pharmacological treatment corresponding to the exact stage of the disease, physiotherapy, and surgery [35]. As a result of medication with L-Dopa, a key element of the typical pharmacological regime for PD, patients are expected to develop side effects such as dyskinesias [43]. Since symptoms vary greatly independently of treatment and PD progresses at different rates in different individuals, treatment requires regular clinical monitoring and medication adjustment.

There are over 130,000 people with Parkinson's in the UK, each individual seen by a specialist doctor or nurse only once or twice a year, allowing only brief and intermittent assessment of the wide range of their motor and non-motor symptoms [37]. This in turn limits opportunities to precisely quantify PD progression and the effectiveness of patient stratification: the restricted availability of data concerning individual variability and actual symptom trends limits opportunities to adapt care to the needs of a particular individual at a specific time. To address this challenge we developed cloudUPDRS, the first smartphone app to achieve certification as a Class I medical device by the Medicines & Healthcare products Regulatory Agency in the UK for the clinical assessment of the motor symptoms of Parkinson's.

cloudUPDRS augments standard clinical care pathways by enabling daily assessments which lead to (i) more consistent and reliable care, (ii) early identification of problems such as medication side-effects, thus enabling earlier intervention, (iii) monitoring of individualised patient trends leading to more effective patient stratification, and (v) enable patients to take ownership of their own care through non-pharmacological measures such as improved nutrition and physical therapy.

The cloudUPDRS system is based on the Universal Parkinson's Disease Rating Scale [14] and the PDQ39 questionnaire [18], and incorporates a cloud-based Big Data management and analytics service to generate objective and reliable assessments of motor performance. Patients use the app at home to record sensor measurements while performing a series of simple actions with each limb, such as tapping the screen to assess bradykinesia and holding the phone on their knee to assess tremor. The data captured by the phone is then used to calculate the clinical UPDRS score through the application of a biomedical signal processing pipeline. Additional longitudinal analytics are performed subsequently to enable trend analysis and patient stratification.

This paper presents two recent developments within cloudUPDRS, specifically:

- A deep learning technique applied to sensor observations so as to assess compliance with the actions dictated by the UPDRS protocol. Combined with a bespoke user experience this technique can replace expert supervision while maintaining high-quality data collection.
- Personalised tests reducing the time required to carry out an assessment to less than 4 minutes. These so-called *quick tests* are created using machine learning to select a subset of UPDRS observations that closely estimate the motor performance of a particular patient.

978-1-5090-4327-9/17/$31.00 ©2017 IEEE

In the following section we review research related to this work and then proceed to report on key factors for patient compliance identified through user research in Section III. We present the cloudUPDRS system in Section IV and in Section V we report on certification. We then proceed to discuss the details of the two techniques identified above in Sections VI and VII correspondingly.

II. Related Work

In [20], we demonstrate the feasibility of using smartphones as a means to assess commonly occurring motor symptoms of PD in a clinical setting. Specifically, we design, develop and validate in a clinical study a prototype app on Android implementing Part III of the MDS-UPDRS[14]. Using the accelerometer and touch screen sensors commonly available in modern smartphones, we are able to carry out hand and leg tremor measurements, as well as gait and bradykinesia assessments using finger tapping tasks to replicate the majority of these tests (cf. Table I and Section IV-A). Other research groups have followed a similar approach focusing on specific symptoms. Most commonly tremor measurements are considered for example [9], [22], [25] and [26] all provide proof-of-concept implementations of upper limb tremor estimation.

Two major projects by the M. J. Fox Foundation in the US currently explore the diversity of PD motor symptoms with a large population sample. The first employs the mPower app (http://parkinsonmpower.org/) by Sage Bionetworks [41] implemented using the open source researchkit framework developed by Apple on iOS (http://researchkit.org/). The focus of this work is to develop a large data set of motor performance observations, which can be used as a benchmark for the experimental evaluation of algorithms providing PD diagnosis. The second project is in collaboration with Intel and The Grove Foundation and employs wearables to provide 24×7 monitoring of PD patients—specifically a Pebble watch (https://www.pebble.com/) to measure wrist tremor relayed via an Android app to a Cloudera-based back end for storage and analysis. The focus of this study is the development of a deep longitudinal data set capturing in minute detail the second-by-second variations of motor symptoms from a population of tenths of thousands of volunteers. However, battery longevity and data transmission issues have limited opportunities to capture complete traces and the project is currently considering alternative strategies.

Recently, the suitability of machine learning has been investigated for the assessment of PD. Voice samples are processed using standard machine learning algorithms in [2] to correlate individual performance and MDS-UPDRS score. A deep learning approach is adopted in [15] to identify patients in ON and OFF states using Restricted Boltzmann Machines to analyse accelerometer data. Both projects report encouraging results which merit further investigation but current performance limitations prevent these techniques from becoming an effective clinical tool.

III. Understanding Patients with PD

The wider adoption of cloudUPDRS by patient communities necessitates that tests are incorporated as part of their daily routine. To understand how to best facilitate this we carried out extensive interviews with clinicians, technologists, patients, carers and patient advocates (22 individuals in total); a web survey with participants from the research volunteer pool of Parkinson's UK receiving 166 unique submissions; and, three audience panels (16 participants in total). Across all studies we recruited participants with a confirmed diagnosis of PD and excluded individuals with generic symptoms of Parkinsonism. Patient participants represent all Hoehn & Yahr levels except for the audience panels in which participation was limited to Level 3, due to the practicalities of access to the venue.[1]

The potentially transformative role of smartphone apps for PD was widely acknowledged in interviews. The expectation of positive outcomes was closely related to recent trends enabling the direct involvement of patients in establishing research priorities, the use of patient expertise in research, and towards greater transparency. This perspective was often related to opportunities for patient empowerment as expressed for example in online communities such as PatientsLikeMe [8], suggesting that evidence-based care must cater for the translation of evidence into practice in a manner directly accessible to and understandable by patients.

We employed the web survey to explore current phone usage patterns specifically among patients and to identify potential constraints that may place barriers for the adoption of the cloudUPDRS app. Responses received were primarily from mobile phone users (96%) with 77% coming from those with a smartphone. The majority of smartphone owners (87%) use it daily with only 14% reporting significant difficulties. A relative small proportion of those with smartphones (20%) use apps to track their symptoms or manage medication. The vast majority (86%) expects to make regular use of the cloudUPDRS app with 64% expressing a preference for the test session lasting a maximum of 5 minutes, 27% accepting a test duration of 10 minutes, and 5% even longer. The majority (68%) expect to make use of the app at least once per day to assess their symptoms.

Audience panels combined elements of user experience evaluation and a wider exploration of perceived costs and benefits of the cloudUPDRS app, which was demonstrated during the sessions. Panelists identified specific problems with the version presented, for example the potential effects of involuntary movements common to specific patient profiles and suggested improvements. As relates to the utility of the app and their motivation for regular use, the opportunity to manage symptoms was an unequivocal benefit for the majority of participants and strongly motivated their involvement. However, access to detailed performance data was less important compared to the sense of understanding afforded

[1] Hoehn & Yahr is a clinical rating scale that defines categories of motor function in PD, ranging from minimal or no functional disability at level 1 to confinement to bed unless aided at Level 5.

978-1-5090-4327-9/17/$31.00 ©2017 IEEE

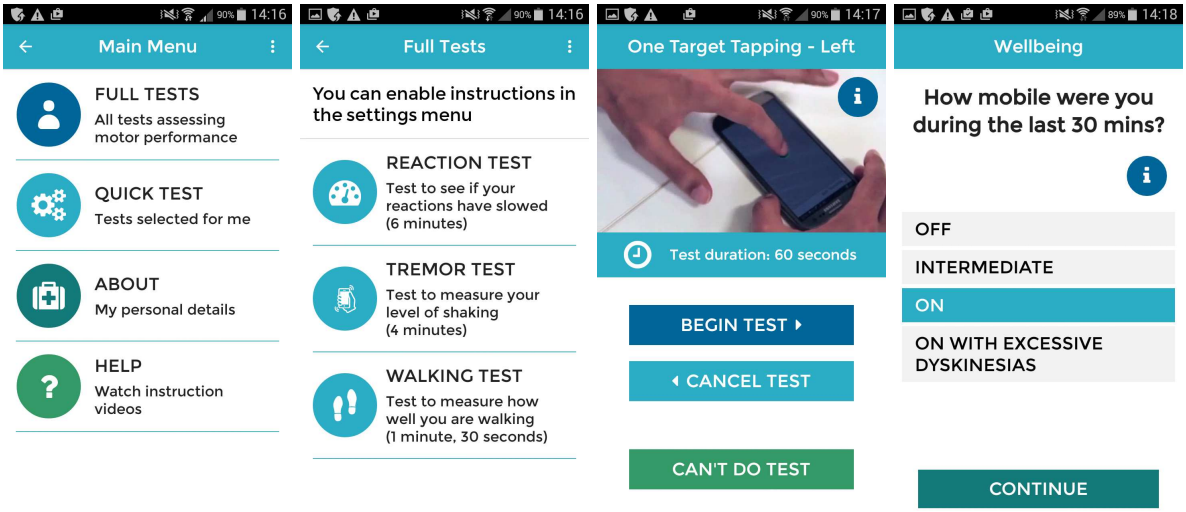

Fig. 1. Views of the user interface of the cloudUPDRS app showing session management, tremor recording and finger tapping activities..

by the experience of using the app and thus of control over the disease. All participants identified with the strong desire to make a contribution towards combating the disease, and considered their donation of personal data and their open availability for research was seen as a means to achieve this goal. As a consequence, no privacy concerns were expressed.

IV. The cloudUPDRS System and App

The complete cloudUPDRS system consists of:

1) A smartphone app for Android that enables patients to carry out motor performance tests and complete a wellness self-assessment; conduct session management; and securely submit data to the cloudUPDRS service.
2) Cloud-based scalable data collection service that ingests data from patients' smartphones; ensures secure data management; and applies the signal processing pipeline.
3) Data-mining toolkit for medical intelligence incorporating quantitative and semi-structured data, and longitudinal analyses, clustering and classification; and a clinical user interface incorporating visualisation.

The cloudUPDRS app implements a comprehensive workflow (partially depicted in Figure 1) that provides audio, video and textual media to guide patients and their carers to conduct the tests at home and in the community unsupervised by a trained clinician or specialist nurse. The app implements a bespoke user interaction design to ensure that the data recorded capture the actual motor performance features as required for the successful application of MDS-UPDRS. Specifically, patients are guided through a carefully orchestrated sequence of actions while the app records sensor measurements. By requiring the execution of specific action sequences the app restricts the degrees of freedom of individual movement and thus imposes structure and disambiguates user context by limiting the range of observed behaviours.[2] As a result, the recorded signal can be interpreted accurately using a small set of heuristics rather than require the use of a full context model and reasoning approach [3]. Finally, the app automatically adapts to match the specifications of its host device and incorporates a delay tolerant service to manage data upload.

The full test administered by the cloudUPDRS app consists of 17 individual observations, specifically kinetic, postural and resting tremor for the left and right hand; left and right leg agility and resting tremor; single and double target finger tapping on both sides; and, gait. During each observation period lasting 60 seconds, the patient is required to assume a specific position and perform the prescribed movement as described in the previous paragraph. Following the recording of these observations the patient is presented with a questionnaire incorporating selected questions from the Parkinson's Disease Questionnaire (PDQ-39) [18] and recording the time of the most recent medication intake.

The cloudUPDRS service is engineered to facilitate scalable performance by adopting the microservices architecture [34]. This approach is set in contrast to traditional monolithic web applications and aims to maximise opportunities for vertical decomposition and scaling-out, which are critical for high performance and service resilience in data intensive situations. cloudUPDRS microservices are implemented as composite Docker containers, are loosely coupled and employ lightweight communication and coordination mechanisms such as the Consumer-Driven Contract pattern. System componentization is enforced via versioning of published RESTful interfaces and sandboxed instances of the service can be deployed automatically to cater for data isolation between distinct regulatory domains. The overall service architecture has been designed for scalability so that real-time streams captured for example during concurrent patient consultations can be integrated on the fly with archival information from the longitudinal datastore service. To facilitate this *modus operandi*, we provide structured workflows implemented through microservices

[2]The action sequences can be seen in video demonstrations available at http://www.updrs.net/help.

978-1-5090-4327-9/17/$31.00 ©2017 IEEE

TABLE I
ANALYTICS TOOLBOX SIGNAL PROCESSING FUNCTIONS AND CORRESPONDENCE TO THE SECTIONS OF THE MDS-UPDRS.

Analytic Function	MDS-UPDRS Section
Rest Tremor	3.17 (rest tremor amplitude)
Postural Tremor	3.15 (postural tremor of the hands)
Action Tremor	3.16 (kinetic tremor of the hands)
Pronation—supination Movements	3.6 (pronation—supination movements of the hands)
Leg agility	3.8 (leg agility)
Finger tapping	3.3 (rigidity) & 3.4 (finger tapping)
Gait	3.10 (gait) & 3.11 (freezing of gait)

following the lambda architecture [30], which facilitates the efficient fusion of real-time and archival data on the fly.

A. Bio-signal Processing

Precise assessment of tremor, bradykinesia and gait is typically carried out using laboratory equipment for example tailor-made biomedical data acquisition systems incorporating transducers such as high-frequency/high-accuracy accelerometers and gyroscopes, signal amplifiers and filters and high-performance analog-to-digital converters. The captured signal is analysed subsequently by specialist commercial software such as Spike 2 by Cambridge Electronic Design Ltd with the total cost of a complete system rising to tenths of thousands.

Laboratory based clinical rating however is constrained by the requirement that the patient is present in the clinic, and in practice can only be carried out as a "snap-shot" assessment. In [20] we show that the sensor, clock and data acquisition hardware of a low-end smartphone captures data with sufficient accuracy to precisely quantify the magnitude of PD motor symptoms across the majority of the tests included in Part III of the MDS-UPDRS by comparing its performance against results obtained using a biomedical analytics system by CED. In cloudUPDRS we automate the methodology presented in [20] as a bespoke cloud-based data analytics service [11]. For completeness of presentation, we briefly summarise the main features of this system here.

1) Tremor: Tremor measurements are recorded for both hands at rest, at posture and in action as listed in Table I. For rest tremor measurements, users are asked to relax their hands on their lap in a supine position while the phone is lying in their palm. For the postural tremor measurements patients are guided to keep their arm outstretched directly on their front while holding the smartphone. Finally, for action tremor measurements they are required to hold the phone and move it between the chest and the fully outstretched position on their front. In all cases, acceleration is recorded along three axes in $\mathrm{m/s^2}$ at the maximum supported sampling rate (at least $50\,\mathrm{Hz}$) and timestamped at maximum resolution (typically microseconds). Tremor is calculated as the cumulative magnitude of the scalar sum acceleration across three axes for all frequencies between $2\,\mathrm{Hz}$ and $10\,\mathrm{Hz}$. To obtain this power spectrum the signal is first filtered with a Butterworth high-pass second order filter at $2\,\mathrm{Hz}$ and the Fast Fourier Transform (FFT) is subsequently applied to the filtered waveform data.

2) Bradykinesia : MDS-UPDRS assess bradykinesia, or else the slowness of movement, through three different factors: (i) pronation-supination movements, (ii) leg agility, and (iii) finger tapping. In the first test patients are asked to hold the phone and perform alternating pronation-supination movements, that is rotating the palm of the hand toward the inside so that it is facing downward and then toward the outside so that the palm is facing upward, as fast and as fully as possible. Leg agility measurements require the phone to be placed on the thigh of the patient while seated, holding the phone lightly with the ipsilateral hand, while raising and stomping the foot on the ground as high and as fast as possible. During both tests the phone is recording acceleration data in a manner similar to the tremor tests. The assessment of the pronation-supination movements and leg agility tests requires the estimation of the frequency and power of movement. To obtain these, the toolkit first removes DC and applies a Butterworth low-pass second order filter at $4\,\mathrm{Hz}$ in order to exclude most of the tremor. Subsequently, the power of the movement is calculated as the total amplitude between $0\,\mathrm{Hz}$ and $4\,\mathrm{Hz}$ and the frequency derived from the power spectrum.

Finger tapping performance is assessed in two tests using single and dual targets presented on the screen of the phone at set locations with patients attempting to tap them as fast and as accurately as possible (alternating between targets in the dual-target case). When tapping accidentally occurs outside the screen area the test is repeated. The touch-sensitive screen of the smartphone is used to collect the information used for performance calculations, specifically the timing of each touch event, its duration, the direction of movement (upwards or downwards), the coordinates on the phone screen, and the amount of pressure applied are recorded. For the two-target variant it is necessary that the distance between targets be at a specific distance irrespective of the size of the screen or of the device. To estimate finger tapping performance the analytical functions first identify all touch events and employ the associated timestamps to estimate tap frequency (taps per second), the mean hand movement time between taps (in milliseconds), and the actual movement distance between alternative tapings in the dual-target case (in centimetres).

3) Gait: MDS-UPDRS assesses gait by considering multiple behaviours including stride amplitude and speed, height of foot lift and heel strike, and turning and arm swing [48]. The cloudUPDRS variant of this test requires the patient to walk along a straight line for five meters, turn around and return to the point of departure, while the smartphone is positioned either in their belt or trousers pocket. Since it is only possible to measure acceleration data from a single point at the waistline we employ the techniques in [27], [28] to estimate stride frequency and length, velocity and turning time.

V. CERTIFICATION

There are numerous wellness and self-tracking apps readily available on all major platforms and many more that have been

978-1-5090-4327-9/17/$31.00 ©2017 IEEE

developed for research. The vast majority of these apps do not conform to the safety, quality, performance and regulatory requirements set for medical devices and as such can only be employed either to encourage a healthy lifestyle or for research purposes correspondingly — but are not tools that can be used to support medical diagnosis. This fact is often explicitly reflected in their terms and conditions of use for example, quoting from a popular Parkinson's Disease app "we cannot, and thus we do not, guarantee or promise that you will personally receive any direct benefits."

Medical devices are regulated and must conform to rules enforced by regional legislation. Within the European Union, harmonisation of regulations across member countries is facilitated by the Medical Devices Directive (MDD), which provide the blueprint for country–specific legislation. Although the MDD considers situations when software would be treated as a medical device it does not explicitly examine smartphone apps and so its provisions are open to interpretation, an issue that we address in this section. Further, the MDD requires that each member state establishes a Competent Authority to provide guidance and enforce regulation of medical devices and in the UK this responsibility lies with the Medicines and Healthcare products Regulatory Agency (MHRA).

Under Article 1 Clause 2(a) of the MDD a medical device is defined as "any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, including the software intended by its manufacturer to be used specifically for diagnostic and/or therapeutic purposes." The current interpretation of this definition by the MHRA as relating to apps implies that "if the [mobile] application is intended to carry out further calculations, enhancements or interpretations of entered/captured patient data, $[\cdots]$ it will be a Medical Device. If it carries out complex calculations, which replaces the clinician's own calculation and which will therefore be relied upon, then it will certainly be considered a Medical Device." Hence, the features of the cloudUPDRS app clearly place it within the provisions of the MDD. For certification purposes, the named publisher of the app on the selected platform store is considered its manufacturer as defined by the MDD, and thus the party obliged to ensure conformity with the provisions of the directive.

According to the MDD, the cloudUPDRS app is considered a Class 1 medical device that is, an active transient non-invasive instrument. Class 1 devices are considered lower risk and as such as less closely regulated. In this case, certification requires that the app meets the Essential Requirements defined in Annex I of the MDD including evidence of software development in compliance with ISO IEC 62304:2006. The app must be supported by comprehensive documentation ensuring that it can be used safely and appropriately by patients and its publisher must "implement and maintain corrective action and vigilance procedures" to ensure safe operation. These requirements add considerable complexity to the development process and in particular require regular review especially when a new version of the app becomes available. cloudUPDRS received medical device status in the UK in May 2016.

VI. Learning Test Movements

In Section IV-A we show how to extend standard lab-based practice for the precise measurement of motor symptoms in PD using a smartphone. In this setting, assessments are supervised by a qualified practitioner who, in addition to helping operate the equipment, can ensure that patients follow closely the actions dictated by Part III of the MDS-UPDRS protocol. However, in the case of self-assessment at home using cloudUPDRS supervision by an expert is unavailable. To address this lack of expert supervision we have designed the guided user experience presented in Section IV, which aims to educate the patient and steer them through the process. While this approach has produced positive results, full compliance with the prescribed actions still cannot be guaranteed or confirmed. Hence, it is necessary that cloudUPDRS provides a mechanism though which data quality can be verified. In particular, it is imperative to introduce a means by which it becomes possible to confirm that the recordings submitted have been captured while the patient performs the required actions correctly[3]. Failure to do so would produce bio-signal measurements that are not representative of the intended tremor type and are likely to result in erroneous scoring.

To achieve this goal, we augment the user experience presented by the app with a deep learning methodology [44]. This approach enables the cloudUPDRS system to learn tremor features associated with a high quality signal and alert the user when an observation has not been captured under satisfactory circumstances. Enabled by recent advances in general-purpose computing using graphics processing units and related algorithmic developments, this methodological approach employs multiple hidden layers to obtain notable results permitting neural networks to identify preferred features directly from the data. This feature of the selected methodology appears especially pertinent to the data quality issue under consideration.

The data set used to investigate the performance of this approach is taken from the first cohort of patients enrolled in the cloudUPDRS trials (8 male and 4 female). Specifically, we consider 227 distinct test sessions conducted over a period of three months (June to August 2016). Data was collected from 9 different phone models providing acceleration measurements at least with a minimum sampling rate of $50\,\mathrm{Hz}$, implemented using the data collection code base of the cloudUPDRS app (other source code elements not affecting data collection were modified during this period). Results are reported specifically for pronation-supination observations of the right hand, without loss of generality for the purposes of this paper. Data captured by the app are normalised but no other pre-processing is performed at this stage.

A. Rationale and Overview

To formulate an algorithmic solution, we reframe the problem of captured data verification as one of binary classification.

[3]In the case of the one- and two-finger tapping tests it is relatively straightforward to identify when the process has been followed accurately directly from the output of the bio-signal processing of Section IV-A.

978-1-5090-4327-9/17/$31.00 ©2017 IEEE

Specifically, the goal of the verification task is to discriminate between high-quality observations and lower-quality sensor recordings captured during movements that do not closely adhere to the guidance of the MDS-UPDRS protocol. To this end, we employ a training data set of observations representing both acceptable and unsuitable cases with known data quality characteristics, guaranteed by the fact that they are collected by the app under controlled conditions or inspected manually. From this data set, features that are distinct within each class are identified algorithmically. Subsequently, the obtained representations are employed to test new observation data submitted by patients via the app: Submissions classified as offering adequate quality are forwarded to the appropriate microservices for data ingestion and signal processing, otherwise they are rejected and excluded from further consideration.

This methodology can be applied asynchronously as part of the data pre-processing and quality assurance phase or interactively, incorporated in the cloudUPDRS app. The latter is possible due to the fact that the classification process has two distinct stages: an initial model training phase representing the most computationally intensive task followed by a sample assessment phase which is relatively lightweight for modern smartphone hardware. As such, the model can be constructed off-line using archival observation data for training and later incorporated in the app, which can conduct real-time quality assessments at the time of data recording and interactively request the repeat of specific individual observations as appropriate to ensure that all submitted tests are usable.

B. Data Segmentation and Pre-processing

Because the duration of each individual observation in cloudUPDRS is 60 seconds and depending on the actual sampling rate supported by the phone used for measurement, the number of samples captured can be relatively very high. For example, sampling at $50\,\mathrm{Hz}$ results to data traces consisting of over $3,000$ records. Rather than incurring the prohibitively excessive computational cost of processing the full sample as a single input we opt to segment the raw signal and consider individual sections separately. For the current investigation and to facilitate manual labelling of the samples we opt to extract the mid-section of the signal. Considering sample sizes of length 256 and 512 respectively, we generate two data sets which we refer to in the following paragaphs as mid-256 and mid-512. As part of this pre-processign step, measurements along the three axes of acceleration expressed in the device coordinate system are supplemented with an extra feature recording the magnitude of acceleration m in Euclidean space.

C. Neural Network Architecture

The core ingredient of our approach is provided by Feed Forward Artificial Neural Networks (FFANN) trained in supervised mode [52]. FFANNs use layers of interconnected neurons represented as matrices of real valued numbers corresponding to connection strengths between the neurons. At each layer neurons perform simple computations, integrating inputs received from neurons at the preceding layer and transforming the signal through activation functions that help regularize data. A geometric interpretation of this process is that activation functions enable FFANNs to partition the high-dimensional data space on which they operate.

This computational model is referred to as the Multilayer Perceptron (MLP) and we will use the term Deep Multilayer Perceptron (DMLP) to refer to the proposed architecture from this point onwards, as it employs five hidden layers whilst the maximum number typically used by simple MLPs is two. In addition, MLPs traditionally employ the $\mathrm{sigmoid}$ and $\tanh$ activation functions because they offer good performance with smaller to medium sized networks. For the DMLP model developed for cloudUPDRS, ReLU or $\mathrm{softplus}$ are preferred instead. This is due to the fact that the latter activation functions can mitigate the vanishing gradient problem which affects deep networks. The function enables them to obtain sparse representations by hard-limiting the input of negative hidden nodes to zero [33].

For learning to happen, the output $\hat{y}$ of the constructed network must be compared against the desired output y, which in the cloudUPDRS case represents the appropriate quality class label that the network should produce, that is, accept or reject the sample. This information is used with a so-called cost or objective function which the DMLP aims to minimise. Here we adopt *categorical cross-entropy* $\mathcal{L}$ as the objective function, defined as $\mathcal{L}(y, \hat{y}) = -y \log \hat{y} - (1 - y) \log(1 - \hat{y})$.

The final step in the process is the application of the backpropagation algorithm [47] which enables the network to learn the distribution that generated the training data. Backpropagation employs the chain rule to calculate the derivatives of the error produced by the objective function with respect to each connection strength between neurons, which are then used to update it. Different versions of the algorithm have been proposed in the literature and in this case we adopt Adam [21], a stochastic optimisation variant.

The standard approach for training MLPs is to feed a single pattern at each step in a stochastic fashion, a process known as Stochastic Gradient Descent (SGD) in the case of the backpropagation algorithm. Although very popular and effective, standard SGD is not preferred with DMLPs mainly due to the prolonged calcualtions required to update the DMLP weights for each pattern in the data set. In cloudUPDRS we adopt instead the mini-batch SGD alternative, where data is fed in small batches and the error averaged out so that only one error signal is propagated for each batch.

To validate the effectiveness of the cloudUPDRS approach we compare its performance against several well-established alternatives selected for their recent success in industrial systems or in highly-regarded competitions such as Kaggle. Full details are provided in Section VI-E below.

D. Classifier Training

Training is carried out separately for each acceleration axis and for the magnitude of acceleration as described in Section VI-B and for each of the mid-256 and mid-512 data sets. Separate training is preferable in this case because it ensures

978-1-5090-4327-9/17/$31.00 ©2017 IEEE

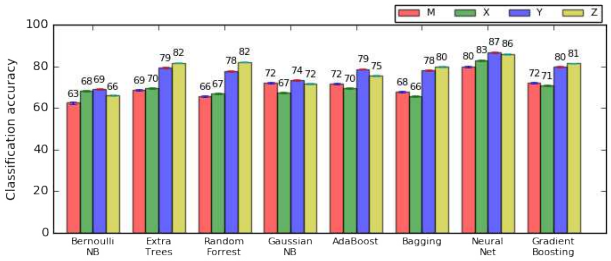

Fig. 2. Average classification success rate for the mid-256 data set.

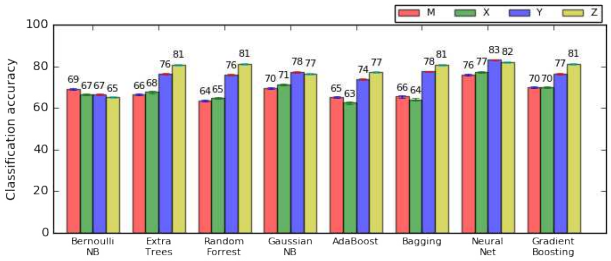

Fig. 3. Average classification success rate for the mid-512 data set.

that each feature is captured accurately and its predictive power can be evaluated independently. Further, the so-called leave-one-out method [1] is combined with early stopping to facilitate the full exploitation of the data set available during training and to reduce the risk of overfitting. The choice of this approach reflects the fact that the data set under investigation has 227 data points which is relatively low in this context.

Each iteration of leave-one-out process involves the exclusion of a single pattern from the full data set, training the classifier on the remaining patterns and testing on the pattern omitted. Consequently the DMLP of Section VI-C is trained as many times as the points available in the data set, in this case 227 times. One limitation of this technique is that it can become biased on the weight initialisation. To address this the process is repeated ten times using different initial random weights and the mean is used as the overall performance metric. Thus, the experiments summarised below are conducted using ten cycles of leave-one-out cross-validation per feature, so that $2,270$ classifiers have been trained and averaged for each of the four features and for each of the data sets.

The early stopping heuristic applied ensures that the learning process is terminated when it reaches a certain predefined threshold. Specifically, we employ three criteria: (i) the categorical cross-entropy or else training error falls below 0.001; (ii) training classification success reaches 100%; or, (iii) the learning process has executed 500 iterations. The benefit of using early stopping is that it prevents the DMLP classifier from memorising counter-productive characteristics discovered in certain samples, especially when these are spurious or irrelevant for the accurate determination of high versus low quality observations. This technique works well when used in conjunction to leave-one-out as it ensures that the DMLP is not over trained [32] on any part of the data set.

E. Results

The deep learning approach described in Sections VI-C and VI-D is implemented using `Keras` (cf. https://keras.io) to provide the description of the DMLP model, on top of the computational graph engine `theano` (cf. http://deeplearning.net/software/theano/). Training was carried out on an array of NVIDIA K40 GPUs achieving a 20-fold speedup against a standard multicore CPU. To provide a baseline against which to evaluate our approach we compare its performance with the following classifiers implemented using the `scikit-learn` [38] machine learning library: (i) Gaussian Naive Bayes [36]; (ii) Bernoulli Naive Bayes [36]; (iii) Random Forest Classifier [7] which employs an ensemble of random decision trees each selected from a sample drawn with replacement; (iv) Extra Trees Classifier [13] is a variation of random forest with thresholds randomly drawn for each candidate feature; (v) AdaBoost Classifier [53] is a meta-estimator which adjusts classifier weights so as to improve learning from difficult classes; (vi) Bagging Classifier [6] is also a meta-estimator which operates on random subsets of the training data to reach a final prediction by aggregating their results; and (vii) Gradient Boosting Classifier [12] which performs optimization of arbitrary differentiable loss functions.

TABLE II
CONFUSION MATRIX FOR DMLP CLASSIFICATION FOR THE MID-256 AND MID-512 DATA SETS (AVERAGED PERFORMANCE). LABELS: T/F TRUE/FALSE AND P/N POSITIVE/NEGATIVE. *A* REPRESENTS THE AVERAGE OF ALL FEATURES.

mid-256 data set

	TP	(%)	FN	(%)	TN	(%)	FP	(%)
M	124.9	88.0	17.1	12.0	56.7	66.7	28.3	33.3
X	142.6	92.6	11.4	7.4	45.5	62.3	27.5	37.7
Y	165.1	95.4	7.9	4.6	32.1	59.4	21.9	40.6
Z	168.2	94.0	10.8	6.0	27.1	56.5	20.9	43.5
A	150.2	**92.5**	11.8	**7.5**	40.4	**61.2**	24.6	**38.8**

mid-512 data set

	TP	(%)	FN	(%)	TN	(%)	FP	(%)
M	122.4	86.2	19.6	13.8	50.1	58.9	34.9	41.1
X	138.1	89.7	15.9	10.3	37.6	51.5	35.4	48.5
Y	164.3	95.0	8.7	5.0	24.5	45.4	29.5	54.6
Z	165.5	92.5	13.5	7.5	21.0	43.8	27.0	56.2
A	147.6	**90.8**	14.4	**9.2**	33.3	**49.9**	31.7	**50.1**

Classification results calculated for all features of the mid-256 data set following the methodology outlined in Section VI-C after training as detailed in Section VI-D are compared to the baseline classifiers in Figure 2. Our methodology employing the DMLP classifier, denoted as Neural Net in the figure, provides the best performance across all features. Note in particular that the magnitude of the standard deviation is very low indicating that initialisation bias has been avoided. Results for the mid-512 data set are calculated following the same approach and summarised in Figure 3. Also in this case, the DMLP approach outperforms all alternative classifiers across all features, often by a significant margin.

Further, the confusion matrix for the DMLP classifier is computed and presented in Table II. Note that the percentage

of false negatives (FN) for the mid-256 data set when all features have been considered is approximately 7% which represents good performance, while false positives (FP) remain below 39%. Hence, the DMLP approach developed for cloudUPDRS identifies correctly the vast majority of low quality samples and causes relatively limited unnecessary repetition of recordings.

VII. Developing the cloudUPDRS Quick Test

In this section, we turn our attention to the development of methods that achieve significant reductions in test duration so as to enable patients to use cloudUPDRS on a daily basis. As suggested by the user studies summarised in Section III, the majority of patients identified a maximum of 5 minutes as the desirable duration for the test. However, even after the initial familiarisation period the full implementation of the procedure typically requires 25 minutes, an estimate that has been confirmed from system logs and independently through user feedback. The critical influence of test duration on user adoption rates was further confirmed during the initial three months of field testing. While the majority of participants carried out tests regularly during the first week following the commission of the app, compliance rates dropped sharply by the end of the third week, and only one out of the 12 participants continued to carry out tests at the end of the three-month testing period.

A. Test Duration and Characteristics

Recall from Section IV, that according to the MDS-UPDRS protocol each individual observation requires 60 seconds of recording and the full test consists of 17 observations, in addition to the medication and well-being questionnaire. Clearly, to reduce the overall duration of the test there are two main options namely to shorten the recording time for individual observations or to reduce the number of observations carried out by selecting a subgroup of the full 17-item set. The final questionnaire requires approximately 30 seconds and is always required because it is used to track medication.

First, consider the option to reduce the length of individual observations without loss of precision in the estimation of motor performance. Specifically, we investigate whether the 60 second observation period set by the MDS-UPDRS protocol is necessary or instead consistent scoring can be still maintained after significantly reducing its duration. To this end, we conduct observations of motor performance for alternative recording periods of 20 and 40 seconds and compare these against measurements carried out for the the full 60 seconds. Tremor and bradykinesia performance metrics were calculated for all observation types in our test data set consisting of 133 full tests carried out by 35 different individuals. In the remainder of this section we report scores calculated for tremor power at rest for the right hand, without loss of generality and so as to specifically quantify our findings.

Figure 4 summarises the results of this analysis and demonstrates that for the majority of patients a shorter observation

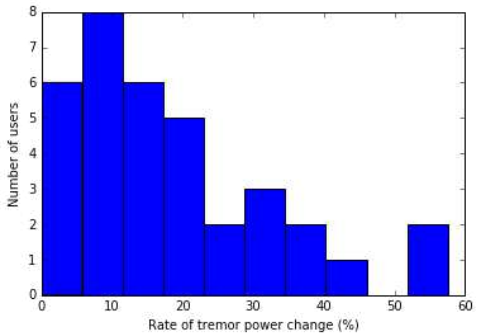

Fig. 4. Change in recorded tremor power between tests of 60 and 20 seconds.

period results in a significant change of their reported motor performance. Specifically, Figure 4 shows that when the recording period is reduced from 60 to 20 seconds the power of the tremor for 60% of the patients is reduced by more than 10%. Similar results are obtained when the observation length is reduced to 40 seconds with the same magnitude of change observed for 35% of the participants in this case. These changes in motor performance for shorter recording periods correspond to significant changes in the estimated MDS-UPDRS score for a single observation ranging between 1 and 2.5 points on the MDS-UPDRS scale. This difference in the actual clinical score corresponds to an average expected disease progression over a six- and twelve-month period respectively, thus representing a significant error in precisely assessing motor performance.

These results clearly imply that that it is necessary to maintain the full 60 second recording period for each individual observation. Relevant clinical literature considering the MDS-UPDRS does not appear to offer explicit justification for this performance. However, it seems that this is an observation readily confirmed by experienced clinicians such as those participating in focus groups conducted by cloudUPDRS (cf. Section III). In particular, it was suggested that the longer duration is required in most cases to cause mild fatigue that reveals the true characteristics of motor performance. In any case, the option to develop the quick test by reducing the duration of individual observations does not appear viable and alternatives must be considered.

B. Identifying Clinically Distinct Factors

Clinical investigations of the MDS-UPDRS scale reported in the medical literature have identified a smaller group of clinically distinct factors, typically five to six, that provide high correlation to the overall score of the motor examination of Part III of the MDS-UPDRS [45], [46]. This observation corroborates the possibility to develop the quick test by reducing the number of individual observations to a much restricted group, which correlates well with the overall patient score. Furthermore, note that the MDS-UPDRS protocol was designed to explore exhaustively the full range of possible motor symptoms caused by PD, but a specific individual would typically present a smaller number of symptoms (especially

978-1-5090-4327-9/17/$31.00 ©2017 IEEE

in earlier stages of PD) that dominate their MDS-UPDRS score and that remain relatively stable over a time frame of a few months. Indeed, a common observation is that PD motor symptoms are asymmetric [5], [40] for example, for a particular patient one side can be significantly affected by tremor while the opposite side may not be affected at all thus contributing zero units towards their MDS-UPDRS score.

Motivated by this observation, we develop a methodology using standard machine learning methods that successfully identify the appropriate subgroup of observations for a specific patient which offer the highest predictive power of their overall motor performance. Upon enrolment in cloudUPDRS, patients are required to carry out the full test at least five times during the first week of monitoring. At the end of this calibration period we use the data of the full test to conduct a feature importance analysis. Specifically, following [13] we apply an ensemble of randomized decision trees on multiple sub-samples of the test data improving its predictive accuracy through averaging and over-fitting control. We then rank individual observations according to the relative importance of their corresponding features (two and three features per tremor and bradykinesia test respectively). Finally, we select the subgroup of top performing observations which account for at least 80% of the variance in the overall UPDRS score.

At the end of this process, the cloudUPDRS system is configured with an individual user profile detailing the subgroup of observations identifed for inclusion in the quick test. This profile is automatically communicated to the app at the next start up so that it is reconfigured to enable the quick test feature in its home screen (cf. Figure 1). The selected settings remain active for a period of six months after which a new set of full tests is required due to the likelihood of changes in motor symptoms over this time frame.

C. Results

To evaluate the effectiveness of this approach we employed the data set described in Section VII-A selecting only patients for which at least five full test results are available. For each patient we apply the above methodology to create a personalised quick test profile. We discover that in all cases we are able to account for the target variance using features associated with only three or less observations. This result is consistent across all patients examined representing medium and progressed stages of the disease.

Figure 5 shows the results of this analysis for a typical patient from this cohort suggesting in this case that just three observations (from which seven features are calculated) are adequate to account for approximately 90% of the variation. Specifically, this patient's quick test profile consists of observations of left leg agility, right arm rest tremor and single tapping of the left hand which provide the adequate information to track their overall motor performance. System logs confirm that this patient was able to complete the quick tests consistently in less than 4 minutes over 50 times in the two months following the availability of their profile. Note that this patient is at an advanced stage of PD presenting significant mobility impairments.

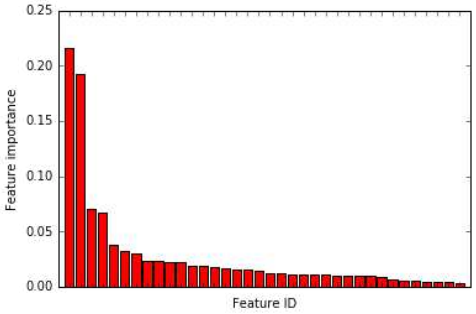

Fig. 5. Predictive power of features associated with individual UPDRS observations.

VIII. Conclusion

According to the World Health Organization [51], ageing populations generate considerable economic effects, notably intensifying pressures on health-care systems which for many of the more economically developed countries already represent the largest area of expenditure. In the UK, the cost of caring for PD patients exceeds 1.25 billion British pounds annually and is rapidly increasing. In this socioeconomic situation, mobile health apps present a unique opportunity for the provision of cost effective care at population scale. Yet, to reach their full potential such apps must offer safety guarantees and facilitate a seamless user experience.

In this paper, we introduced two novel techniques developed for cloudUPDRS, a medical device app for the assessment of motor symptoms of PD at home, addressing these requirements. First, a bespoke deep learning approach was employed to replace expert human supervision of the administration of the common motor performance assessment protocol for PD; and second, a personalised quick test was developed to accurately trace overall motor performance while considerably improving patient compliance. In our experiments both approaches performed reliably and produced promising results. We anticipate both techniques to be useful for a wider class of mobile health-care apps with similar requirements. Further experimentation with a larger patient population is of course necessary to fully assess the potential of the two techniques developed and we are currently working towards this within the CUSSP clinical study.

Acknowledgments

Project cloudUPDRS: Big Data Analytics for Parkinson's Disease patient stratification is supported by Innovate UK (Project Number 102160). We gratefully acknowledge the support of NVIDIA Corporation with the donation of Tesla K40 GPUs used for this research. Project partners would also like to thank Parkinson's UK for providing access to their online forums and assisting with the recruitment of participants.

REFERENCES

[1] D. M. Allen. "The relationship between variable selection and data augmentation and a method for prediction." *Technometrics*, 16(1), 125-127, 1974.

[2] S. Arora, V. Venkataraman, A. Zhang, S. Donohue, K.M. Biglan, E.R. Dorsey and M.A. Little. "Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study." *Parkinsonism and Related Disorders*, 21(6), 650–653, 2015.

[3] C. Bettini, O. Brdiczka, K. Henricksen, J. Indulska, D. Nicklas, A. Ranganathan and D. Riboni. "A survey of context modelling and reasoning techniques." *Pervasive and Mobile Computing*, 6(2), 161–180, 2010.

[4] J. Bergstra, Y. Bengio. "Random Search for Hyper-parameter Optimization." *J. of Machine Learning Research*, 13, 281-305, 2012.

[5] O. Blin, A. M. Ferrandez and G. Serratrice. "Quantitative analysis of gait in Parkinson patients: increased variability of stride length." *J. Neurol. Sci.*, 98, 91–97, 1990.

[6] L. Breiman. "Bagging Predictors." *Mach. Learn.*, 24(2), 123–140, 1996.

[7] L. Breiman. "Random Forests." *Mach. Learn.*, 45(1), 5–32, 2001.

[8] J.R. Brubaker, C. Lustig and G.R. Hayes. "PatientsLikeMe: empowerment and representation in a patient-centred social network." *CSCW10 Workshop Research in Healthcare: Past, Present, and Future*, Savannah, GA, USA, 2010.

[9] J.F. Daneault *et al.*. "Using a smart phone as a standalone platform for detection and monitoring of pathological tremors." *Front Hum Neurosci*, 6, 357, 2008.

[10] European Brain Council. *Parkinson's disease Fact Sheet*, 2011.

[11] N.F. Fragopanagos, S. Kueppers, P. Kassavetis, M.U. Luchini, and G. Roussos. "Towards Longitudinal Data Analytics in Parkinson's Disease." *Proc. 1st Int. Conf. on IoT and Big Data Technologies for HealthCare*, June 15-16, Budapest, Hungary, 2016.

[12] J. Friedman. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics*, 29(5), 1189–1232, 2001.

[13] P. Geurts, D. Ernst and L. Wehenkel. "Extremely Randomized Trees." *Mach. Learn.*, 63(1), 3–42, 2006.

[14] C.G. Goetz *et al.*. "Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale Presentation and Clinimetric Testing Results." *Movement Disorders*, 23(15), 2129–2170, 2008.

[15] N. Y. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker and T. Ploetz. ' 'PD Disease State Assessment in Naturalistic Environments Using Deep Learning." *AAAI Conf. on Artificial Intelligence*, 2015.

[16] S. Hochreiter and J. Schmidhuber. "Long Short-Term Memory." *Neural Computation*, 9(8), 1735–1780, 1997.

[17] J. Jankovic. "Parkinsons disease: clinical features and diagnosis." *J. Neurology, Neurosurgery and Psychiatry*, 79(4), 368-76, 2008.

[18] C. Jenkinson, R. Fitzpatrick, V. Peto, R. Greenhall and N. Hyma. "The Parkinson's Disease Questionnaire (PDQ-39): development and validation of a Parkinson's disease summary index score." *Age Ageing*, 26(5), 353–357, 1997.

[19] A. Jha, P. Kassavetis, E Nomikou, J. Rothwell, K. Bhatia and G. Roussos. "The cloudUPDRS smartphone app: home monitoring for Parkinson's Disease." *The Future of Medicine Conference*, Royal Society of Medicine, May 19, London, UK, 2016.

[20] P. Kassavetis, T. A. Saifee, G. Roussos, L. Drougkas, M. Kojovic, J. C. Rothwell, M. J. Edwards, K. P. Bhatia. "Developing a tool for remote digital assessment of Parkinson's disease." *Movement Disorders - Clinical Practice*, 3(1), 2015.

[21] D. Kingma and J. Ba. "Adam: A method for stochastic optimization." *arXiv preprint* arXiv:1412.6980, 2015.

[22] N. Kostikis *et al.*. "Towards remote evaluation of movement disorders via smartphones." *Proc. IEEE Eng Med Biol Soc*, 5240–3, 2011.

[23] S. Kueppers, I. Daskalopoulos, A. Jha, N.F. Fragopanagos, P. Kassavetis, E. Nomikou, T. Saifee, J.C. Rothwell, K. Bhatia, M.U. Luchini, M. Iannone, T. Moussouri, and G. Roussos. "From Wellness to Medical Diagnostic apps: The Parkinson's Disease Case." *Proc. Int. Conf. on Personal, Pervasive and mobile Health*, June 14-16, Budapest, 2016.

[24] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel. "Back-propagation applied to handwritten zip code recognition." *Neural Computation*, 1(4):541–551, 1989.

[25] R. Lemoyne *et al.* "Implementation of an iPhone for characterizing Parkinson's disease tremor through a wireless accelerometer application." *Pro.c IEEE Eng. Med. Biol. Soc.*, 4954–8, 2010.

[26] W. Maetzler, J. Domingos, K. Srulijes, J. J.Ferreira and B. R. Bloem. "Quantitative wearable sensors for objective assessment of Parkinson's disease,' *Movement Disorders*, 28(12), 1628–1637, 2013.

[27] E. Martin. "Novel method for stride length estimation with body area network accelerometers." *IEEE BioWireleSS*, 79–82, 2011.

[28] E. Martin, V. Shia and R. Bajcsy. "Determination of a Patient's Speed and Stride Length Minimizing Hardware Requirements." *Proc. Int. Conf. Body Sensor Networks*, 144–149, 2011.

[29] V. Marx. "Human phenotyping on a population scale." *Nature Methods*, 12, 711–714, 2015.

[30] N. Marz and J. Warren. *Big Data: Principles and best practices of scalable realtime data systems*. Manning Publications, 2013.

[31] P.M. Matthews, P. Edison, O.C. Geraghty and M.R. Johnson. "The emerging agenda of stratified medicine in neurology." *Nature Reviews*, 10, 15–27, 2014.

[32] J. Moody. "Prediction Risk and Architecture Selection for Neural Networks." *From Statistics to Neural Networks: Theory and Pattern Recognition Applications*, 136, 147–165, 1994.

[33] V. Nair and G. E. Hinton. "Rectified linear units improve restricted Boltzmann machines." *Int. Conf. Machine Learning (ICML)*, 807–814, 2010.

[34] S. Newman. *Building Microservices: Designing Fine-Grained Systems*. O'Reilly Media, 2015.

[35] National Institute for Health and Clinical Excellence. *Parkinson's disease: diagnosis and management in primary and secondary care: National cost-impact report*. NICE clinical guideline no. 35, 2006.

[36] A. McCallum and K. Nigam. "A comparison of event models for naive Bayes text classification." *Proc. AAAI/ICML-98 Work. on Learning for Text Categorization*, 41–48, 1998.

[37] Parkinson's UK. *Parkinson's prevalence in the United Kingdom*. http://www.parkinsons.org.uk/sites/default/files/parkinsonsprevalenceuk_0.pdf, 2009.

[38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. "Scikit-learn: Machine learning in Python." *J. of Machine Learning Research*, 12, 2825–2830, 2011.

[39] L. Y. Pratt. "Non-literal transfer of information among inductive learners." In R. J. Mammone and Y. Y. Zeevi (ed.) *Neural Networks: Theory and Applications II*, 1992.

[40] L. Ricciardi, D. Ricciardi, F. Lena *et al.*. "Working on asymmetry in Parkinson's disease: randomized, controlled pilot study." *Neurol. Sci.*, 36, 1337–1343, 2015.

[41] R. Robinson. "Electronic Sensors Break New Ground in Neurology Practice and Research," *Neurology Today*, 15(7), 20–26, 2015.

[42] A. Rodriguez-Molinero *et al.*. "Validation of a portable device for mapping motor and gait disturbances in Parkinson's disease." *JMIR Mhealth Uhealth*, 3(1), e9, 2015.

[43] A. H. V. Schapira, M. Emre, P. Jenner and W. Poewe. "Levodopa in the treatment of Parkinson's disease." *European Journal of Neurology*, 16, 982–989, 2009.

[44] J. Schmidhuber. "Deep learning in neural networks: An overview." *Neural Networks*, 61, 85–117, 2015.

[45] G. T. Stebbins and C. G. Goetz. "Factor structure of the unified Parkinson's disease rating scale: Motor examination section." *Movement Disorders*, 13(4), 633–636, 1998.

[46] S. D. Vassar *et al.*. "Confirmatory Factor Analysis of the Motor Unified Parkinson's Disease Rating Scale." *Parkinson's Disease*, Article ID 719167, 2012.

[47] P. Werbos. *Beyond regression: new tools for prediction and analysis in the behavioral sciences*. Ph.D. Thesis, Harvard University, 1974.

[48] M. W. Whittle. *Gait Analysis: An introduction*. Butterworth-Heinemann, 2014.

[49] P. Wicks. *The patient of the future*. http://parkinsonsmovement.com/the-patient-of-the-future/, 2015.

[50] D. H. Wolpert. "Stacked Generalization." *Neural Net.*, 5, 241–259, 1992.

[51] World Health Organization. *Current Status of the World Health Survey*. http://www.who.int/healthinfo/survey/. 2013.

[52] G. P. Zhang. "Neural networks for classification: a survey." *IEEE Trans. Systems, Man, and Cybernetics, Part C*, 30(4), 451–462, 2000.

[53] J. Zhu, H. Zou, S. Rosset, T. Hastie. "Multi-class AdaBoost." *Stat. and Interface*, 2, 349–360, 2009.

978-1-5090-4327-9/17/$31.00 ©2017 IEEE