The Journal of Engineering

Research Article

Proposal of a health care network based on big data analytics for PDs

eISSN 2051-3305
Received on 13th July 2018
Accepted on 7th January 2019
doi: 10.1049/joe.2018.5142
www.ietdl.org

Leonarda Carnimeo[1], Gianpaolo Francesco Trotta[2], Antonio Brunetti[1], Giacomo Donato Cascarano[1], Domenico Buongiorno[1], Claudio Loconsole[1], Eugenio Di Sciascio[1], Vitoantonio Bevilacqua[1] ✉

[1]*Department of Electrical and Information Engineering (DEI), Polytechnic University of Bari, Bari (BA), Italy*

[2]*Department of Mechanics, Mathematics and Management Engineering (DMMM), Polytechnic University of Bari, Bari (BA), Italy*

✉ *E-mail: vitoantonio.bevilacqua@poliba.it*

Abstract: Health care networks for Parkinson's disease (PD) already exist and have been already proposed in the literature, but most of them are not able to analyse the vast volume of data generated from medical examinations and collected and organised in a pre-defined manner. In this work, the authors propose a novel health care network based on big data analytics for PD. The main goal of the proposed architecture is to support clinicians in the objective assessment of the typical PD motor issues and alterations. The proposed health care network has the ability to retrieve a vast volume of acquired heterogeneous data from a Data warehouse and train an ensemble SVM to classify and rate the motor severity of a PD patient. Once the network is trained, it will be able to analyse the data collected during motor examinations of a PD patient and generate a diagnostic report on the basis of the previously acquired knowledge. Such a diagnostic report represents a tool both to monitor the follow up of the disease for each patient and give robust advice about the severity of the disease to clinicians.

1 Introduction

Parkinson's disease (PD) is currently one of the most spread neurodegenerative disorders. In detail, it is a degenerative brain disorder characterised by a loss of midbrain dopamine neurones [1, 2] and the main clinical PD symptoms related to body movements involve tremor, rigidity, bradykinesia, and gait abnormalities. Unfortunately, no definitive treatment is at the moment available. Nevertheless, it has been proved that the quality of life of patients can be increased by means of other novel therapies. Physician evaluations are commonly based on historical information from the patient, regarding motor function during activities of daily living and clinic observation, using clinical rating scales and the number of patients is growing as well as the amount of medical data [3]. On this proposal, health care networks for PD already exist [4], but the vast volume of existing medical data leaves most of them still unanalysed.

More in detail, related health care systems generally consist of a variety of providers (e.g. medical centres, hospital departments, and emergency rooms) and prescribers (e.g. general and specialist physicians). Relationships exist among these entities, but an entire network should be assessed. Several works address the problem of analysing medical data by applying data/process mining and simulation [5–7] and in machine learning and deep learning applied in the medical field [8–33].

Nowadays, health care organisations involving both single-physician offices with multi-provider groups and large hospital networks with accountable care organisations stand to realise significant benefits by using big data for effectively digitising or combining them [34, 35]. It seems that existing analytical techniques can be applied to the vast amount of existing (but currently unanalysed) patient-related medical data to reach a deeper understanding of outcomes to be applied at the point of care. Potential benefits could include following up specific diseases as well as Parkinson's pathologies. In general, big data analytics (BDA) in health care could contribute to evidence-based medicine for analysing a lot of structured and unstructured medical data to match treatments with outcomes, device/remote monitoring for capturing, in real time, large volumes of fast-moving data from several devices placed at home or in hospital and, finally, patient profile analytics for applying several analysis to patient profile to improve cares and lifestyles.

In this paper, it is suggested a health care network based on BDA for PD to support clinicians in the objective assessment of the typical PD motor issues and alterations by means of an ensemble support vector machine (ESVM). In particular, the network architecture presented in this paper aims to suggest a solution which avoids the situation, in which a patient has to go to his health care, as shown in Fig. 1*a*. For this purpose, a health care network based on a big data system (BDS) is proposed and properly described, as shown in Fig. 1*b*.

The considered BDS should be composed of four main elements:

I. Big data sources.
II. Big data transformation.
III. Big data tools and platforms.
IV. BDA for PD.

As shown in Fig. 2, volume, velocity, variety, and veracity are features of a BDS and, in particular, they are features of a BDS in health care [35].

In detail, volume: A BDS creates and accumulates over time an incredible amount of health-related data such as personal, medical records, radiology images, clinical trial data, three-dimensional (3D) imaging, genomics, and biometric sensor readings. Biometric data from sensors and other types of common data such as clinicians' notes, video, and images are considered in this work. Advances in data management, particularly virtualisation and cloud computing are currently facilitating the development of platforms for more effective captures, storages, and manipulations of large volumes of data [36].

Velocity: Data are accumulated and analysed in real time and at a rapid pace or velocity. In many medical situations, the application of these features in a BDS could do the difference between life and death. For example, in this work, the ability to retrieve, analyse, compare, and make decisions based on output values could help physicians to have a diagnostic report in a brief period of time, thanks to the MapReduce ESVM (MRESVM) approach used to predict the disease severity of patients.

Variety: One of the things that make big data really big is that they are coming from several different sources. The recent exploiting of these sources for analytics means that so-called structured data (which previously held unchallenged hegemony in

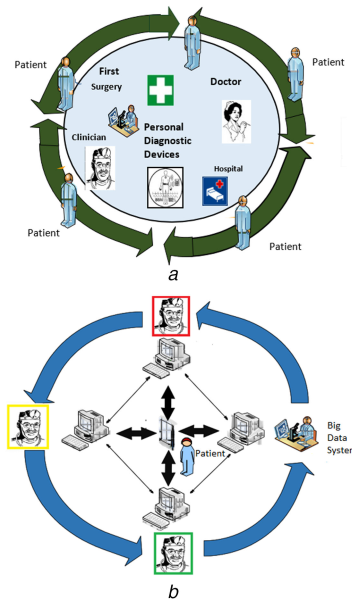


Fig. 1 *Network architecture*
(a) Traditional health care network, ***(b)*** Proposed health care network based on a BDS

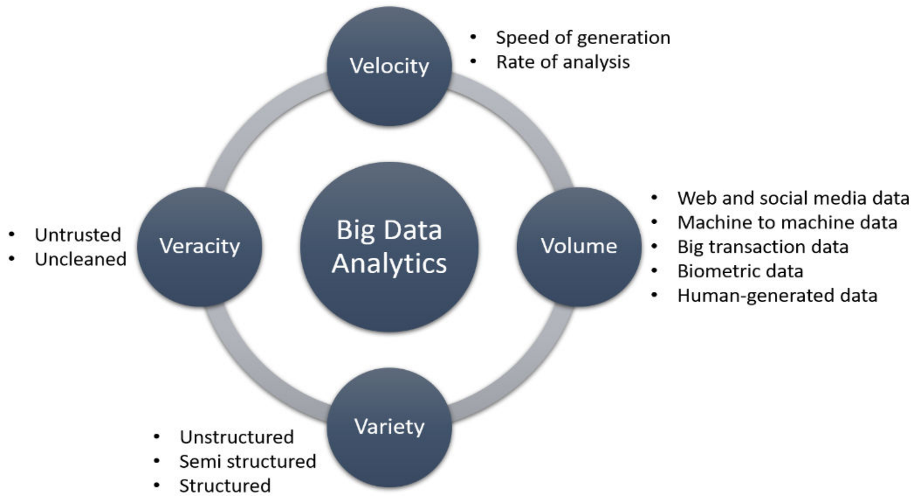


Fig. 2 *Four 'Vs' of BDA in health care*

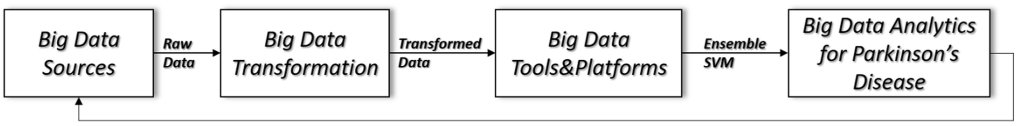


Fig. 3 *Block diagram of the considered BDS*

analytics) is now joined by unstructured data (text and human language) and semi-structured data (XML, RSS feeds). There is also data that is hard to categorise, as it comes from audio, video, and other devices. Moreover, multidimensional data can be drawn from a Data warehouse to add historical context to big data. It is a far more eclectic mixture of data types than analytics has ever involved. So, with big data, variety is just as big as volume. In addition, variety and volume tend to fuel each other [37]. In this work, a lot of data from several sensors and application are stored and processed.

Veracity or 'data assurance': The quality of health care data is highly variable, and life or death decisions depend on having accurate information. Unstructured data imply all often incorrect. The veracity hypothesises a scaling up in the performance of techniques and technologies to use in a big data management system. BDA in health care could be executed across several servers, or nodes, in distributed processing and by considering the use of the paradigm of parallel computing and of the approach called 'divide and process'. Moreover, models and techniques need to take into account the characteristics of BDA. Traditional data management hypothesises the warehoused data is certain, precise, and clean. High-quality data enable improving coordination of care, avoiding errors, and reducing costs. Due to the fact that BDA has to be exactly specified for each field of application, a detailed BDS suitable for PD is herein proposed and reported in Fig. 3. On the basis of this premise, in this paper, a proposal of a health care network based on BDA for PD is described as reported in Fig. 4.

More in detail, the big data source component consists of three main sub-components (patient, clinicians' team for PD, and the acquisition system). The interaction between these sub-components has several heterogeneous data including images, videos, electromyographic data, kinetic and kinematic data, clinicians' notes, patient personal data etc. as outcomes. In particular, the acquisition system is designed and developed to acquire and evaluate several features both from a motor impairment examination and a handwriting analysis. The big data

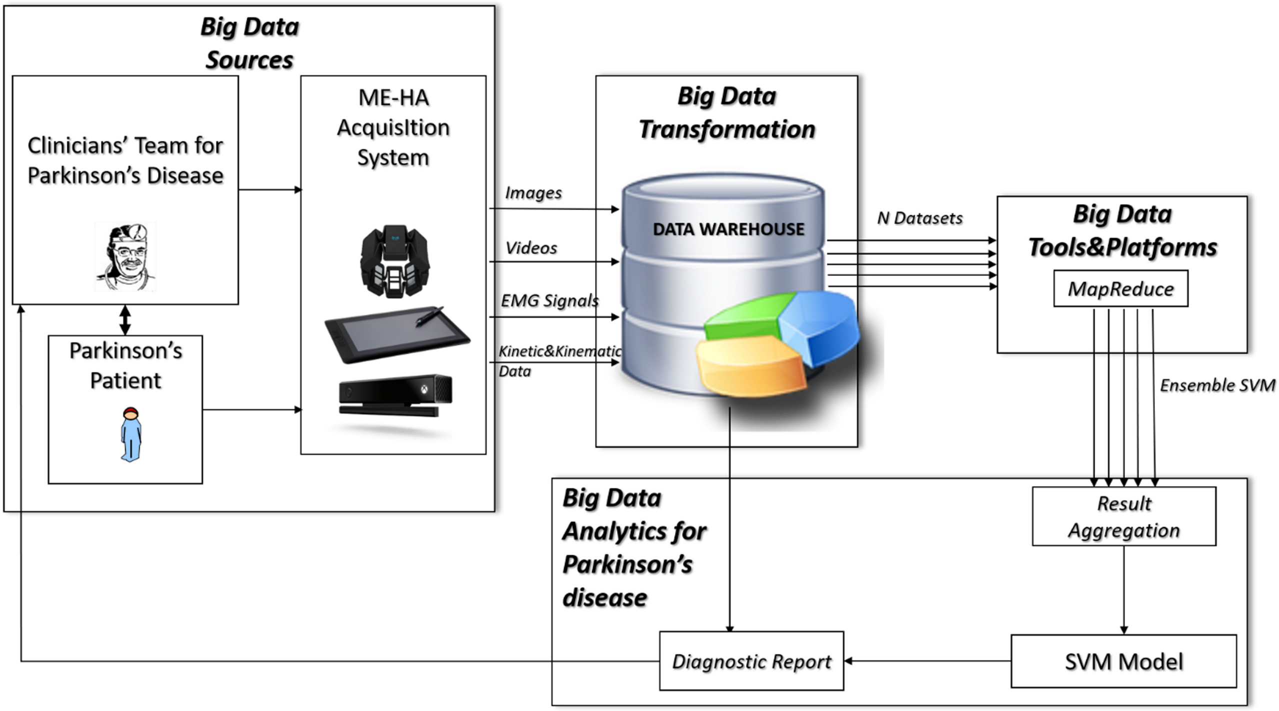


Fig. 4 *Block diagram of the proposed health care network based on BDA for PD*

transformation component consists of a Data warehouse that could store structured, semi-structured, and unstructured data that come from the first component. In particular, since unstructured data is random and difficult to analyse, structured data, already tagged and easily sorted, and semi-structured data that contains tags to separate data elements [38], are stored in a Data warehouse. The Big Data Tools & Platforms is a component that uses structured or semi-structured data from the Big Data Transformation component as input. In this work, this component performs the implementation of an ESVM [39] based on several datasets collected in a proper Data warehouse. The BDA for PD component is used to perform a result aggregation to describe the severity of the PD so that each patient can be monitored during the disease progression. Finally, a diagnostic report based on the output of the SVM model generated from the ESVM is a feedback to clinicians.

2 Big data sources

As shown in Fig. 4, sources and data types generally involve [40]:

i. web and social media data;
ii. machine-to-machine data;
iii. big transaction data;
iv. biometric data; and
v. human-generated data.

(i) It is related to clickstream and interaction data from Facebook, Twitter, LinkedIn, and blogs like health plan websites and smartphone apps. (ii) It concerned data read from remote sensors, meters, and other vital sign devices. (iii) It is the type of data types and sources involves health care claims and other billing records increasingly available both in semi-structured and unstructured formats. (iv) It is related to fingerprints, genetics, handwriting, retinal scans, and other medical images, blood pressure, and other similar types of data. Finally, (v) it is the type of data types and sources involves unstructured and semi-structured data such as physician's notes, paper documents, and Email.

2.1 Clinicians' team for PD

Nowadays, there are several foundations that lead to the development of new treatments and care models through its network. In particular, the National Parkinson Foundation consists of 42 leading medical centres around the world. These centres of excellence:

- Deliver care to more than 100,000 people with PD.
- Represent a community of health care professionals dedicated to improve the lives of everyone with PD today.
- Represent exemplary care and research to improve the lives of their own patients, people in the community who do not have the benefit of centre of excellence care, and for future generations.

In this paper, the procedures are described on how to obtain a rigorous modelling and a meaningful analysis of the large volume of data produced by a health care network based on BDA for PD. This analysis can be useful in a clinician's network for Parkinson's patients. All final outcomes of this network can be an improvement in the quality of service perceived by patients and in the following up of a Parkinson's patient. The proposed health care network based on BDA for PD is organised into four main parts as shown in Fig. 4:

- in 'Big Data Sources', the interactions between clinicians' network, patients, and a novel acquisition system for motor examination and handwriting analysis useful to help clinicians to monitor the follow up of the disease are proposed;
- in 'Big Data Transformation', inputs come from the data collected by the clinicians' team for PD and are clustered in datasets;
- in 'Big Data Tools & Platforms', an MR algorithm is used to create an ESVM to classify data; and
- in 'BDA for PD', the SVM model is created and a diagnostic report consisting of the output of the SVM model and the data collected in Data warehouse not used to train the ESVM is generated.

The conclusions are that the obtained results can be useful for health care managers to inform them clearly about the status of services under their responsibility, and to suggest improvements to system inefficiencies. It is also useful to evaluate the degree of collaboration among different entities of the network.

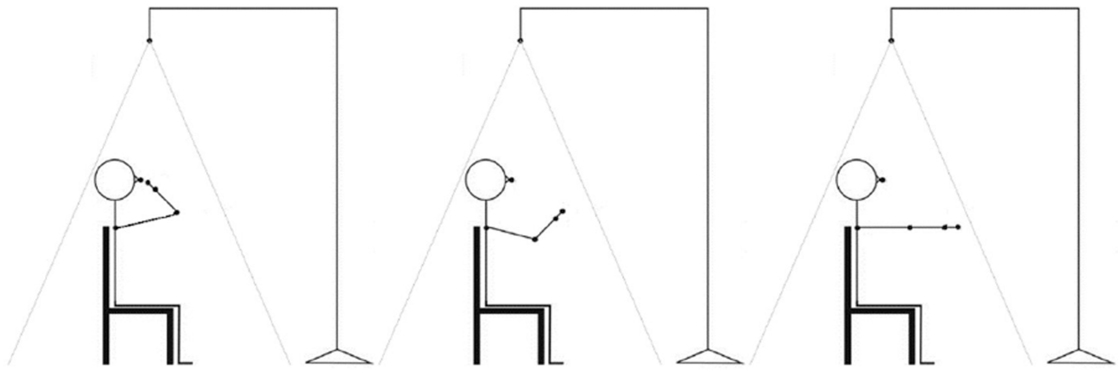


Fig. 5 *Example of the exercise execution*

2.2 Acquisition system for motor examination and handwriting analysis

The motor exercises considered in the acquisition system are a subset of exercises described in the Unified PD Rating Scale (UPDRS) reviewed by the Movement Disorder Society (MDS) in 2007 [41] and that consists of four parts: part I (non-motor experience in daily living), part II (motor experience in daily living), part III (motor examination), and part IV (motor complications). Nowadays, there are several scientific results supporting its validity. Subjectivity and low efficiency are inevitable as most of the diagnostic criteria use descriptive symptoms, which cannot provide a quantified diagnostic basis. In particular, main problems regard the evaluation of the severity of specific symptoms such as freezing of gait [42–44], dysarthria [45], tremor [46–49], bradykinesia [36, 50, 51], and dyskinesia [40, 52–56]. Another interesting research field focuses on the analysis of different common life tasks such as handwriting, which is a highly overlearned fine and complex manual skill involving an intricate blend of cognitive, sensory, and perceptual-motor components [57]. For these reasons, the presence of an abnormality in the handwriting process is a well known and well-recognised manifestation of a wide variety of neuromotor diseases. There are two main difficulties related to handwriting which affect PD patients: (i) the difficulty in controlling the amplitude of the movement, i.e. a decreased letter size (micrography) and failing in maintaining stroke width of the characters as writing progresses [58] and (ii) the irregular and bradykinetic movements, i.e. increased movement time, decreased velocities and accelerations, and irregular velocity and acceleration trends over time [59]. For these reasons, in the literature, there are several works investigating the possibility of differentiation between PD patients and healthy subjects by means of computer-aided handwriting analysis tools [60, 61]. By considering the current state of the art, important novel contributions are described here with designing and evaluating two specific systems for PD patients: a vision-based system able to capture specific movements of different main MDS-UPDRS [1, 62–65] scale exercises and a handwriting analysis tool able to extract biometric signals related both to pen movements and muscular activity. Furthermore, a specific set of features extracted from the previous system set-up is evaluated. The motor examination-handwriting analysis (ME-HA) acquisition system consists of several instruments:

- *Microsoft Kinect*: Attached to a telescopic bar along the vertical axis to enable users to change its orientation and position so that it can recognise movements [22].
- Passive reflective markers to track the position of fingers and toes.
- *Myo armband*: A wearable gesture control and motion control device consisting of eight electromyographic sensors used as surface Electromyography (sEMG) bracelet sensor for acquiring sEMG signals from eight different points of the forearm.
- WACOM Cintiq 13’ HD used as graphics tablet providing co-located visual feedback to acquire pen tip position (planar *x–y* coordinates) and pressure, and the tilt of the pen with respect to the writing surface.

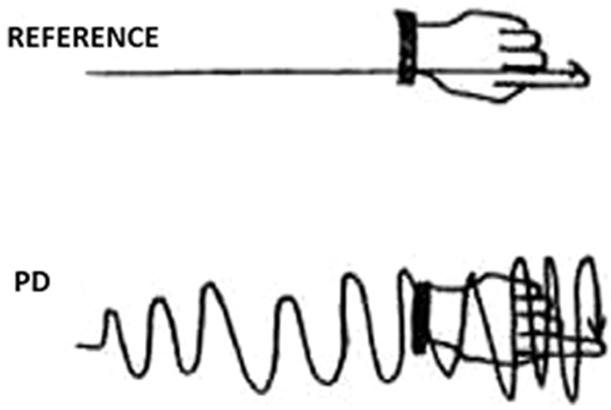


Fig. 6 *Example of trajectory difference between PD and ideal one*

Several features based on all tasks performed by each patient have to be derived; thus, proper acquisition system needs to be designed and developed by means of these instruments. In particular, several systems based on the third part of MDS-UPDRS regarding the motor examination and the handwriting analysis can be designed.

Finger-to-nose task [49]*:* The Kinect® sensor is placed at a height of 80 cm above the patient's head. So, from a raised position, it can capture a patient's location and the hand involved in the experiment. In addition, Kinect® has been preferred to normal red, green, and blue (RGB) cameras because it captures 3D information, using a depth sensor. It consists of an infrared camera, which involves realising calculations using a different wavelength than the traditional RGB camera. Finally, it is a relatively low-cost technology, which makes the whole system extremely cheap. The kinetic tremor is evaluated which measures the smooth, coordinated movement of the upper limbs by having each of the examinees touch the tip of their nose with their index finger (Fig. 5).

In every examination, all features required for the classification have been evaluated by means of a reflector marker placed on a finger of each patient. An example of the typical trajectory during the exercise for PD and control is depicted in Fig. 6.

Medical specialists usually observe significant features useful for evaluating tremor only on the basis of their own experience, i.e. by adopting a qualitative and subjective approach in their evaluation. Some of these features useful for evaluating tremor include average amplitude, maximum amplitude, frequency, and are subsequently determined and analysed by this system. A cloud of points, each one representing the spatial position of the marker over three axes (*X*, *Y*, and *Z*) is acquired during the finger-to-nose experiment; this cloud is extracted from images captured at a frame rate of 30 fps. Proper algorithms are subsequently used to extract the features of interest from the cloud of points as described in [49].

Finger tapping and foot tapping tasks [63, 66, 67]*:* This could be developed in two separate vision-based systems able to acquire the movement of the thumb, the index finger, and the toes. Both acquisition systems are based on passive markers made of reflective material and the Microsoft Kinect RGB depth camera (Fig. 7).

The finger tapping exercise set-up considers the examination of both hands separately. While the subject is seated in front of the camera, he has to tap ten times the index finger on the thumb quickly. During the task, the subject wears two thimbles made of reflective material on both the index finger and thumb. Meanwhile, in the foot tapping exercise set-up, the feet are tested separately. The tested subject sits on a straight-backed chair in front of the camera and has both feet on the floor. He is then instructed to place the heel on the ground in a comfortable position and then tap the toes ten times as big and as fast as possible. A system of stripes with reflective material is positioned on the toes. The two vision-based acquisition systems use passive reflective markers to track the position of the thumb, the index finger, and the toes. After the movement acquisition, an image processing phase is needed to recognise the marker in each acquired video frame and compute the 3D position of a centroid point associated with the specific marker. This post-processing phase has been conducted using the OpenCV library running the following steps on each image frame:

- conversion to a grey-scale image;
- extraction of the pixels associated with the reflective passive markers with a thresholding operation;
- blurring and thresholding operations in sequence;
- eroding and dilating operations in sequence; and
- dilating and eroding operations in sequence.

After the post-processing phase, all the found blobs are extracted using an edge detection procedure. Only the blobs having sizes comparable with markers' size are kept for the next analysis. As a final step, the centroid of each blob (only one blob for the foot tapping and two blobs for the finger tapping) is computed. Given the position of the centroid, its depth information and the intrinsic parameters of the used camera, have computed the 3D position of the centroid associated with each tracked marker in the camera reference system. Such centroid has then considered as the position of the specific finger or of the foot's toes.

The entire post-processing analysis described above produces the 3D positions of toes' marker (foot tapping) and of the two fingers' markers (finger tapping). Given the position of each marker, we then extracted the following signals over time:

- $d1(t)$: The distance between the two fingers' markers over time in finger tapping (Fig. 8*a*).
- $d2(t)$: The distance between the position of the toes' marker over time and the position of the same marker when the toes are completely on the ground in foot tapping (Fig. 8*b*).

Given the entire acquired signal, all the single trials have been extracted for each side. The set of the extracted features contains features of the time domain, space domain, and frequency domain as described in [63].

Handwriting analysis [56, 58, 59, 61]*:* This technique is based on a model-free technique that allows the extraction and the classification of particular features starting from the assumption that the characteristics (or features) of one or more particular biometric signals or parameters can synthesise and represent a particular aspect of the user's handwriting. The analysis requires the application of processing algorithms on signals generated starting from a specifically created pattern and can succeed to extract the features of interest.

The exercise, in particular, requires the writing of specific pattern and the recording of two main different sources of information. The system set-up is reported in Fig. 9 and includes the MyoTM Gesture Control Armband (www.myo.com) used as sEMG bracelet sensor for acquiring sEMG signals from eight different points of the forearm and the WACOM Cintiq 13″' HD (www.wacom.com/en-ch/products/pen-displays/cintiq-13-hd) used as graphics tablet providing co-located visual feedback to acquire pen tip position (planar *x*–*y* coordinates) and pressure, and the tilt of the pen with respect to the writing surface.

To perform the analysis, three writing patterns, corresponding to as many writing tasks, have been selected. They were properly differentiated according to the writing size and as the size constraint (one task is size unconstrained, while the other two have a visual reference as for size constraint):

- *Task number 1*: A five turn spiral drawn in an anti-clockwise direction.
- *Task number 2*: A sequence of eight 'l' with a size of 2.5 cm (a 2.5 cm visual marker reference is displayed on the left of the tablet screen).
- *Task number 3* A sequence of eight 'l' with a size of 5 cm (a 5 cm visual marker reference is displayed on the left of the tablet screen).

Several features have been extracted. The acquisition set-up allows the synchronous recording of two main sources of information and signals representing different aspects of handwriting: the sEMG of the forearm and the pen data from the tablet. The algorithms used to evaluate features are described in [56, 60, 61].

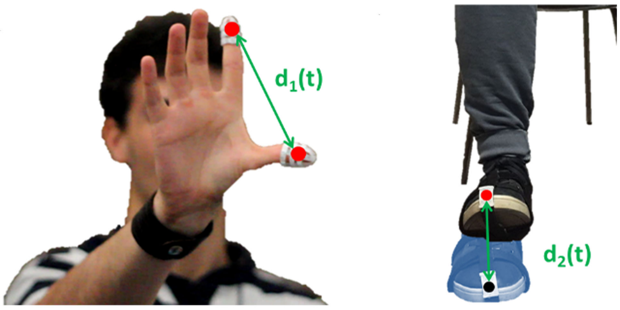
Gait analysis [44, 64]*:* Motion analysis data collection started with the subject standing in a T-pose for one second to facilitate the skeleton tracking. Subjects then walked toward the Kinect sensor, which was placed 3.5 m away from the subject's starting point at a height of 0.75 m. The 3.5 m distance was selected to guarantee that the recorded gait cycle, which began when the subject was about 2.5–3 m from the Kinect, did not include the acceleration/deceleration phases of walking that are anticipated during the initiation or completion of the gait task (Fig. 10).

Three categories of features have been considered for the gait analysis as described in [57]:

- *Temporal:* To assess the duration of gait phases in seconds and in percentage compared with the duration of the gait cycle (stance and swing phase/time, double support phase, and stride time).
- *Spatial:* To estimate length, width, and velocity of movements, normalised by the height or the lower limb length of the subject (stride cadence/length/velocity, step length/width, and swing velocity).
- *Angular:* To assess the degree of rotation for specific postures and movements, typical of Parkinsonian patients (trunk/neck flexion, Pisa syndrome, and arm swing).

At the end of an examination, in which the patient performs all the tasks and the clinician acquires all the data necessary for the classification, a large volume of data is generated and stored in a Data warehouse. In particular, a big training dataset is created and consists of features evaluated during the examinations by means of the devices of ME-HA system acquisition or post-processed by means of algorithms designed and developed as described in the works cited below for each task. However, these features are not the only data generated from the framework created to help the clinician to perform the medical examination. In fact, for the creation of a dataset of a supervised classification algorithm, a label for each entry is necessary. In this case, the label for each task performed by each patient is the disease severity. To insert this type of data in a Data warehouse, at the end of the exercise, the graphical user interface, used by the clinician to use instruments to acquire movements during the examination, is also used to label each exercise performed by each patient. Moreover, remotely, a different clinician, belonging to clinicians' team for PD could label an exercise; thus making the result of the examination more objective. In fact, during an examination, not only the features are extracted, but also a video of the examination is saved. This is useful to allow other clinicians to review the examination. In this way, a dataset of clinical results is created and should be used to label the dataset of features extracted and generated during the examination. Obviously, other types of data should be saved in a Data warehouse and should be used for the classification process (age, gender etc.) or for the creation of the diagnostic report (video, images, the plot of the movements trend during the examination etc.).

Fig. 7 *Left image shows a healthy subject wearing the two passive finger markers. The images reported on the right-hand side show the foot of a subject doing the foot tapping exercise while he is wearing a passive marker on the toes*



a *b*

Fig. 8 *Finger tapping and foot tapping*
(a) Finger tapping: the signal $d1(t)$ is the distance between the two centroids (red-filled circles) of the passive finger markers, ***(b)*** Foot tapping: the signal $d2(t)$ is the distance between the centroid of the toes' marker and the centroid of the same

3 Big data transformation

The heterogeneous data from the Big Data Sources component of the proposed model should be stored in the Big Data Transformation component, which consists of a Data warehouse able to store several types of data. In particular, the data types considered in this paper and used to create datasets and to monitor the follow up of the disease are:

- *Video:* Acquired during the motor examinations, are useful for reviewing the exercises of a patient also by clinicians that are not present during the real execution. So, this data type is useful to label datasets starting from the knowledge of a network of domain experts.
- *Images:* Acquired during the handwriting analysis, are used to evaluate and to extract some features described below. Moreover, considering the progress of the machine learning algorithms, it will be possible to use these data to extract some new features or to use it as the input of a deep or machine learning network.
- *EMG signals:* Acquired during the handwriting analysis, are signals that describe the trends of muscles activation during the task. A patient with PD is affected by muscles spasms and so, these data, are used to study the frequency and the intensity of these events.
- *Kinetic and kinematic signals:* Acquired during the motor examinations are signals that described the trends of particular points in the space during the task. From these data, it is possible to evaluate several features useful to classify the impairment as is described below.
- *Clinicians notes:* Acquired during an examination, are useful information allowing other clinicians or the SVM experts, to calibrate datasets based on other pathologies of which the patient is affected.
- *Clinical results:* Acquired at the end of the examination, by means of a graphical user interface that helps all the clinicians belonging to clinician's team for PD, to insert the severity of the disease for each task performed by each patient. This step is necessary for the creation of the dataset to use for training a net by means of a supervised algorithm.

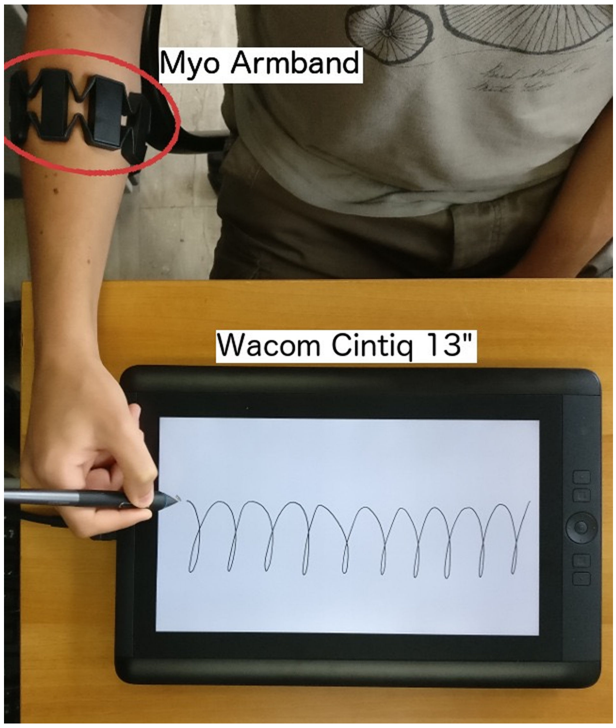


Fig. 9 *System set-up used for the experimental tests to validate the proposed approach*

Fig. 10 *Representation of the proposed set-up in the clinical centre. The patient is asked to walk toward the camera*

To cluster and analyse, in a proper manner, the data stored in the Data warehouse, a data transformation process should be applied. For this scope, the open-source Apache Foundation project named Hadoop should be used to analyse data and to create N datasets used as input in an ESVM. Hadoop has two primary components, namely HaDoop File System (HDFS) and MR programming framework. The main feature of Hadoop is that HDFS and MR are co-deploying such that a single cluster is produced and the storage system is included in the processing system [68].

4 Big data tools and platforms

In the literature, there are several examples to avoid the risk of misclassification by using an ensemble of classifiers (ESVM) in many applications. In [69], Kim *et al.* explain how a bagged ESVM could perform better than a simple SVM classifier, despite the training process is more computationally intensive at the increasing of the dataset dimension [68, 70]. It is well known that a good strategy consists of an aggregation of different numbers of SVM classifiers able to perform good results also for Parkinson's severity task. An ensemble of classifiers is a set of multiple classifiers combining a number of weak learners to create a strong

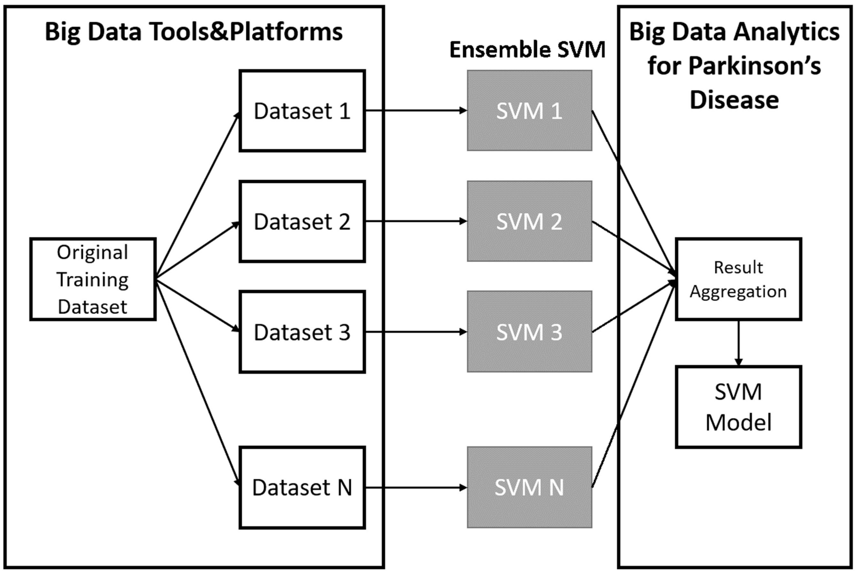


Fig. 11 *Representation of the proposed set-up in the clinical centre. The patient is asked to walk toward the camera*

learner. To create different classifiers, it is possible to use the bootstrap aggregation (bagging) and then increase the variability of the training sets and at the same time preserve the performance in terms of reliability of the final prediction.

To process a dataset with a huge number of entries in a parallel and distributed way, the Google File System [71], that is, an example of implementation of MR programming model, provides an excellent environment. The steps of a general MR model are:

- iterate over the input;
- compute key/value pairs for each input;
- group all intermediate values by key;
- iterate over the resulting groups; and
- reduce each group.

As shown in Fig. 11 that represents an MR framework that should be used in the proposed model, in the map task (or mapper) individual input records are processed in parallel. Then, the system shuffles and sorts the outputs and sends them to the reduce step. Therefore, in the reducer, all associated records are processed by a single entity. Finally, the number of maps and reduce tasks could be defined by users.

5 BDA for PD

In this component, several decisions generally concern with the data input approach to involve as well as preferred distributed design, tool selection, and analytics models. It is well known that the four typical applications of BDA in health care include queries, reports, on-line analytical processing (OLAP), and data mining, whereas visualisation is a common component for each application. Several technologies and techniques have been designed, developed, and evaluated to aggregate, manipulate, analyse, and visualise big data in health care starting from fields as statistics, economics, computer science, and applied mathematics.

In detail, for this proposed health care network based on BDA for PD, the MRESVM algorithm described in [72] consists of several steps. First, from the initial training dataset, a bootstrap algorithm creates *m* training datasets. Then, each training dataset has to be allocated to a single map task. After the allocation, the training of an SVM for each map task is performed and, from each map task, the output is a set of support vectors. A combination of the support vectors is necessary to create a training dataset for the second layer and, finally, the output of the second layer is the trained model.

The trained model is then used to classify the disease severity of a patient that performs the motor examinations and the handwriting analysis over time. The results have to be intended as important support for clinicians' diagnosis, and not all a substitution to a medical decision. Finally, a diagnostic report, consisting of the SVM model results and several data acquired during examinations collected in a Data warehouse, is generated. The report could be useful to collect the results of the examination and to create a historical overview of the patient disease over time.

6 Conclusion and future works

The proposed model of a health care network based on BDA for PD should be capable to transform the way health care providers use sophisticated technologies to gain insight from their clinical and other data repositories and make satisfying decisions. Moreover, the use of machine learning algorithms contextualised in the proposed BDS can support clinicians in their decisions and in monitoring the follow up of each patient's disease. In the future, attention will be focused on issues such as guaranteeing privacy, safeguarding security, establishing standards, and governance, thus improving tools and technologies always more. The authors have already designed and developed the ME-HA acquisition system described in this work and, in future, the proposed health care network will be more accurately implemented, tested, and validated.

7 References

[1] Goetz, C.G.: 'The unified Parkinson's disease rating scale (UPDRS): status and recommendations', *Mov. Disord.*, 2003, **18**, (7), pp. 738–750, doi: 10.1002/mds.10473

[2] Tysnes, O.B., Storstein, A.: 'Epidemiology of Parkinson's disease', *J. Neural Transm.*, 2017, **124**, (8), pp. 901–905. cited By 52, doi: 10.1007/s00702-017-1686-y

[3] Wong, S.L., Gilmour, H.L., Ramage-Morin, P.L.: '*Parkinson's disease: prevalence, diagnosis and impact*' (Statistics Canada, Ottawa, 2014)

[4] Keus, S., Oude-Nijhuis, L., Nijkrake, M., *et al.*: 'Improving community health care for patients with Parkinson's disease: the Dutch model', *Parkinson's Dis.*, 2012, **2012**, Article ID 543426

[5] Mans, R.S., Schonenberg, M.H., Song, M., *et al.*: 'Application of process mining in health care – a case study in a Dutch hospital', *Commun. Comput. Inf. Sci.*, 2008, **25 CCIS**, pp. 425–438, doi: 10.1007/978-3-540-92219-3_32

[6] Di Leva, A., Femiano, S.: 'The BP-M* methodology for process analysis in the health sector', *Intell. Inf. Manage.*, 2011, **3**, (2), p. 56

[7] Antonelli, D., Bruno, G.: 'Health care network modeling and analysis', *IFIP. Adv. Inf. Commun. Technol.*, 2014, **434**, pp. 691–698

[8] Bevilacqua, V., Salatino, A.A., Di Leo, C., *et al.*: 'Advanced classification of Alzheimer's disease and healthy subjects based on EEG markers'. 2015 Int. Joint Conf. Neural Networks IJCNN 2015, Killarney, Ireland, 12–17 July 2015, pp. 1–5, doi: 10.1109/IJCNN.2015.7280463

[9] Buongiorno, D., Barsotti, M., Sotgiu, E., *et al.*: 'A neuromusculoskeletal model of the human upper limb for a myoelectric exoskeleton control using a reduced number of muscles'. 2015 IEEE World Haptics Conf. (WHC), Evanston, USA, June 2015, pp. 273–279, doi: 10.1109/WHC.2015.7177725

[10] Carnimeo, L., Nitti, R.: 'On classifying diabetic patients' with proliferative retinopathies via a radial basis probabilistic neural network', in Huang, D.S., Han, K. (Eds.): '*Advanced intelligent computing theories and applications*' (Springer International Publishing, Cham, 2015), pp. 115–126

[11] Bevilacqua, V., Carnimeo, L., Mastronardi, G., *et al.*: 'On the comparison of NN-based architectures for diabetic damage detection in retinal images', *J. Circuits Syst. Comput.*, 2009, **18**, (8), pp. 1369–1380, doi: 10.1142/S0218126609005721
[12] Carnimeo, L., Altomare, A.R., Nitti, R.: 'A combined preprocessing method for retinal vessel detection towards proliferative diabetic retinopathy screening', in Rossi, F., Mavelli, F., Stano, P., *et al.* (Eds.): '*Advances in artificial life, evolutionary computation and systems chemistry*' (Springer International Publishing, Cham, 2016), pp. 106–116
[13] Bevilacqua, V., Brunetti, A., Cascarano, G.D., *et al.*: 'A deep learning approach for the automatic detection and segmentation in autosomal dominant polycystic kidney disease based on magnetic resonance images', in Huang, D.S., Jo, K.H., Zhang, X.L. (Eds.): '*Intelligent computing theories and application*' (Springer International Publishing, Cham, 2018), pp. 643–649, doi: 10.1007/978-3-319-95933-7_73
[14] Bevilacqua, V., Buongiorno, D., Carlucci, P., *et al.*: 'A supervised CAD to support telemedicine in hematology'. 2015 Int. Joint Conf. Neural Networks (IJCNN), Killarney, Ireland, 2015, pp. 1–7, doi: 10.1109/IJCNN.2015.7280464
[15] Bevilacqua, V., Brunetti, A., Trotta, G.F., *et al.*: in Esposito, A., Faundez-Zanuy, M., Morabito, F.C., *et al.* (Eds.): '*A novel deep learning approach in haematology for classification of leucocytes*' (Springer International Publishing, Cham, 2019), pp. 265–274, doi: 10.1007/978-3-319-95095-2_25
[16] Brunetti, A., Buongiorno, D., Trotta, G.F., *et al.*: 'Computer vision and deep learning techniques for pedestrian detection and tracking: a survey', *Neurocomputing*, 2018, **300**, pp. 17–33, doi: https://doi.org/10.1016/j.neucom.2018.01.092
[17] Bevilacqua, V., Dimauro, G., Marino, F., *et al.*: 'A novel approach to evaluate blood parameters using computer vision techniques'. 2016 IEEE Int. Symp. Medical Measurements and Applications (MeMeA), Benevento, Italy, 2016, pp. 1–6, doi: 10.1109/MeMeA.2016.7533760
[18] de Tommaso, M., Ricci, K., Delussi, M., *et al.*: 'Testing a novel method for improving wayfinding by means of a P3b virtual reality visual paradigm in normal aging', *SpringerPlus*, 2016, **5**, (1), p. 1297, doi: 10.1186/s40064-016-2978-7
[19] Bevilacqua, V., Simeone, S., Brunetti, A., *et al.*: 'A computer aided ophthalmic diagnosis system based on tomographic features', in Huang, D.S., Hussain, A., Han, K., *et al.* (Eds.): '*Intelligent computing methodologies*' (Springer International Publishing, Cham, 2017), pp. 598–609, doi: 10.1007/978-3-319-63315-2_52
[20] Bevilacqua, V., Altini, D., Bruni, M., *et al.*: 'A supervised breast lesion images classification from tomosynthesis technique', in Huang, D.S., Jo, K.H., Figueroa-García, J.C. (Eds.): '*Intelligent computing theories and application*' (Springer International Publishing, Cham, 2017), pp. 483–489, doi: 10.1007/978-3-319-63312-1_42
[21] Bevilacqua, V., Mastronardi, G., Piscopo, G.: 'Evolutionary approach to inverse planning in coplanar radiotherapy', *Image Vis. Comput.*, 2007, **25**, (2), pp. 196–203, doi: 10.1016/j.imavis.2006.01.027
[22] Manghisi, V.M., Uva, A.E., Fiorentino, M., *et al.*: 'Real time RULA assessment using Kinect v2 sensor', *Appl. Ergon.*, 2017, **65**, pp. 481–491, doi: 10.1016/j.apergo.2017.02.015
[23] Menolascina, F., Bellomo, D., Maiwald, T., *et al.*: 'Developing optimal input design strategies in cancer systems biology with applications to microfluidic device engineering', *BMC Bioinf.*, 2009, **10**, (Suppl. 12), S4, doi: 10.1186/1471-2105-10-S12-S4
[24] Buongiorno, D., Barone, F., Solazzi, M., *et al.*: 'A linear optimization procedure for an EMG-driven neuromusculoskeletal model parameters adjusting: validation through a myoelectric exoskeleton control', in Bello, F., Kajimoto, H., Visell, Y. (Eds.): '*Haptics: perception, devices, control, and applications*' (Springer International Publishing, Cham, 2016), pp. 218–227, doi: 10.1007/978-3-319-42324-1_22
[25] Buongiorno, D., Barone, F., Berger, D.J., *et al.*: 'Evaluation of a pose-shared synergy-based isometric model for hand force estimation: towards myocontrol', in Ibáñez, J., González-Vargas, J., Azorín, J.M., *et al.* (Eds.): '*Converging clinical and engineering research on neurorehabilitation II*' (Springer International Publishing, Cham, 2017), pp. 953–958, doi: 10.1007/978-3-319-46669-9_154
[26] Buongiorno, D., Barsotti, M., Barone, F., *et al.*: 'A linear approach to optimize an EMG-driven neuromusculoskeletal model for movement intention detection in myo-control: a case study on shoulder and elbow joints', *Front. Neurorobot.*, 2018, **12**, p. 74, doi: 10.3389/fnbot.2018.00074
[27] Menolascina, F., Tommasi, S., Paradiso, A., *et al.*: 'Novel data mining techniques in a CGH based breast cancer subtypes profiling: the biological perspective'. 2007 IEEE Symp. Computational Intelligence and Bioinformatics and Computational Biology, Honolulu, Hawaii, USA, 2007, pp. 9–16, doi: 10.1109/CIBCB.2007.4221198
[28] Bevilacqua, V., Tattoli, G., Buongiorno, D., *et al.*: 'A novel BCI-SSVEP based approach for control of walking in virtual environment using a convolutional neural network'. 2014 Int. Joint Conf. Neural Networks (IJCNN), Beijing, China, 2014, pp. 4121–4128, doi: 10.1109/IJCNN.2014.6889955
[29] Bevilacqua, V., Cariello, L., Columbo, D., *et al.*: 'Retinal fundus biometric analysis for personal identifications', in Huang, D.S., Wunsch, D.C., Levine, D.S., *et al.* (Eds.): '*Advanced intelligent computing theories and applications. With aspects of artificial intelligence*' (Springer, Berlin, Heidelberg, 2008), pp. 1229–1237, doi: 10.1007/978-3-540-85984-0_147
[30] Brunetti, A., Carnimeo, L., Trotta, G.F., *et al.*: 'Computer-assisted frameworks for classification of liver, breast and blood neoplasias via neural networks: a survey based on medical images', *Neurocomputing*, 2019, **335**, pp. 274–298, doi: https://doi.org/10.1016/j.neucom.2018.06.080
[31] Bevilacqua, V., Brunetti, A., Trotta, G.F., *et al.*: 'A deep learning approach for hepatocellular carcinoma grading', *Int. J. Comput. Vis. Image Process.*, 2017, **7**, (2), pp. 1–18, doi: 10.4018/IJCVIP.2017040101
[32] Carnimeo, L., Bevilacqua, V., Cariello, L., *et al.*: 'Retinal vessel extraction by a combined neural network–wavelet enhancement method', in Huang, D.S., Jo, K.H., Lee, H.H., *et al.* (Eds.): '*Emerging intelligent computing technology and applications. With aspects of artificial intelligence*' (Springer, Berlin, Heidelberg, 2009), pp. 1106–1116, doi: 10.1007/978-3-642-04020-7_118
[33] Bortone, I., Trotta, G.F., Cascarano, G.D., *et al.*: 'A supervised approach to classify the status of bone mineral density in post-menopausal women through static and dynamic baropodometry'. 2018 Int. Joint Conf. Neural Networks (IJCNN), Rio de Janeiro, Brazil, 2018, pp. 1–7, doi: 10.1109/IJCNN.2018.8489205
[34] Burghard, C.: 'Big data and analytics key to accountable care success', *IDC Health Insights*, 2012, pp. 1–9
[35] Raghupathi, W., Raghupathi, V.: 'Big data analytics in health care: promise and potential', *Health. Inf. Sci. Syst.*, 2014, **2**, (1), p. 3, doi: 10.1186/2047-2501-2-3
[36] Griffiths, R.I., Kotschet, K., Arfon, S., *et al.*: 'Automated assessment of bradykinesia and dyskinesia in Parkinson's disease', *J. Parkinson's Dis.*, 2012, **2**, (1), pp. 47–55, doi: 10.3233/JPD-2012-11071
[37] Rajaraman, V.: 'Big data analytics', *Resonance*, 2016, **21**, (8), pp. 695–716, doi: 10.1007/s12045-016-0376-7
[38] Sagiroglu, S., Sinanc, D.: 'Big data: a review'. 2013 Int. Conf. Collaboration Technologies and Systems (CTS), San Diego, USA, 2013, pp. 42–47, doi: 10.1109/CTS.2013.6567202
[39] Priyadarshini, A.: 'A map reduce based support vector machine for big data classification', *Int. J. Database Theory Appl.*, 2015, **8**, (5), pp. 77–98, doi: 10.14257/ijdta.2015.8.5.07
[40] Liu, X., Carroll, C.B., Wang, S.Y., *et al.*: 'Quantifying drug-induced dyskinesias in the arms using digitised spiral-drawing tasks', *J. Neurosci. Methods*, 2005, **144**, (1), pp. 47–52, doi: https://doi.org/10.1016/j.jneumeth.2004.10.005
[41] Twelves, D., Perkins, K.S., Counsell, C.: 'Systematic review of incidence studies of Parkinson's disease', *Mov. Disord.*, 2003, **18**, (1), pp. 19–31
[42] Djurić-Jovičić, M.D., Jovičić, N.S., Radovanović, S.M., *et al.*: 'Automatic identification and classification of freezing of gait episodes in Parkinson's disease patients', *IEEE Trans. Neural Syst. Rehabil. Eng.*, 2014, **22**, (3), pp. 685–694, doi: 10.1109/TNSRE.2013.2287241
[43] Tripoliti, E.E., Tzallas, A.T., Tsipouras, M.G., *et al.*: 'Automatic detection of freezing of gait events in patients with Parkinson's disease', *Comput. Methods Programs Biomed.*, 2013, **110**, (1), pp. 12–26, doi: https://doi.org/10.1016/j.cmpb.2012.10.016
[44] Bortone, I., Trotta, G.F., Brunetti, A., *et al.*: 'A novel approach in combination of 3D gait analysis data for aiding clinical decision-making in patients with Parkinson's disease', in Huang, D.S., Jo, K.H., Figueroa-García, J.C. (Eds.): '*Intelligent computing theories and application*' (Springer International Publishing, Cham, 2017), pp. 504–514, doi: 10.1007/978-3-319-63312-1_44
[45] Tsanas, A., Little, M.A., McSharry, P.E., *et al.*: 'Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests', *IEEE Trans. Biomed. Eng.*, 2010, **57**, (4), pp. 884–893, doi: 10.1109/TBME.2009.2036000
[46] Mellone, S., Palmerini, L., Cappello, A., *et al.*: 'Hilbert–Huang-based tremor removal to assess postural properties from accelerometers', *IEEE Trans. Biomed. Eng.*, 2011, **58**, (6), pp. 1752–1761, doi: 10.1109/TBME.2011.2116017
[47] Heldman, D.A., Espay, A.J., LeWitt, P.A., *et al.*: 'Clinician versus machine: reliability and responsiveness of motor endpoints in Parkinson's disease', *Parkinsonism Relat. Disord.*, 2014, **20**, (6), pp. 590–595, doi: https://doi.org/10.1016/j.parkreldis.2014.02.022
[48] Rigas, G., Tzallas, A.T., Tsipouras, M.G., *et al.*: 'Assessment of tremor activity in the Parkinson's disease using a set of wearable sensors', *IEEE Trans. Inf. Technol. Biomed.*, 2012, **16**, (3), pp. 478–487, doi: 10.1109/TITB.2011.2182616
[49] Bevilacqua, V., Trotta, G.F., Loconsole, C., *et al.*: 'A RGB-D sensor based tool for assessment and rating of movement disorders', in Duffy, V., Lightner, N. (Eds.): '*Advances in human factors and ergonomics in health care and medical devices*' (Springer International Publishing, Cham, 2018), pp. 110–118, doi: 10.1007/978-3-319-60483-1_12
[50] Salarian, A., Russmann, H., Wider, C., *et al.*: 'Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system', *IEEE Trans. Biomed. Eng.*, 2007, **54**, (2), pp. 313–322, doi: 10.1109/TBME.2006.886670
[51] Dai, H., Zhang, P., Lueth, T.C.: 'Quantitative assessment of Parkinsonian tremor based on an inertial measurement unit', *Sensors*, 2015, **15**, (10), pp. 25055–25071, doi: 10.3390/s151025055
[52] Keijsers, N.L., Horstink, M.W., Gielen, S.C.: 'Automatic assessment of levodopa-induced dyskinesias in daily life by neural networks', *Mov. Disord.*, 2003, **18**, (1), pp. 70–80
[53] Lopane, G., Mellone, S., Chiari, L., *et al.*: 'Dyskinesia detection and monitoring by a single sensor in patients with Parkinson's disease', *Mov. Disord.*, 2015, **30**, (9), pp. 1267–1271
[54] Saunders-Pullman, R., Derby, C., Stanley, K., *et al.*: 'Validity of spiral analysis in early Parkinson's disease', *Mov. Disord.*, 2008, **23**, (4), pp. 531–537
[55] Westin, J., Ghiamati, S., Memedi, M., *et al.*: 'A successful computer method for assessing drawing impairment in Parkinson's disease: P2633', *Eur. J. Neurol.*, 2009, **16**, p. 559
[56] Loconsole, C., Trotta, G.F., Brunetti, A., *et al.*: 'Computer vision and EMG-based handwriting analysis for classification in Parkinson's disease', in Huang, D.S., Jo, K.H., Figueroa-García, J.C. (Eds.): '*Intelligent computing

theories and application' (Springer International Publishing, Cham, 2017), pp. 493–503, doi: 10.1007/978-3-319-63312-1_43

[57] Carmeli, E., Patish, H., Coleman, R.: 'The aging hand', *J. Gerontol. A*, 2003, **58**, (2), pp. M146–M152, doi: 10.1093/gerona/58.2.M146

[58] Van Gemmert, A.W.A., Teulings, H.L., Contreras-Vidal, J.L., *et al.*: 'Parkinson's disease and the control of size and speed in handwriting', *Neuropsychologia*, 1999, **37**, (6), pp. 685–694, doi: https://doi.org/10.1016/S0028-3932(98)00122-5

[59] Drotár, P., Mekyska, J., Smékal, Z., *et al.*: 'Prediction potential of different handwriting tasks for diagnosis of Parkinson's'. 2013 E-Health and Bioengineering Conf. (EHB), Iasi, Romania, 2013, pp. 1–4, doi: 10.1109/EHB.2013.6707378

[60] Loconsole, C., Cascarano, G.D., Brunetti, A., *et al.*: 'A model-free technique based on computer vision and sEMG for classification in Parkinson's disease by using computer-assisted handwriting analysis', *Pattern Recognit. Lett.*, 2019, **121**, pp. 28–36, doi: https://doi.org/10.1016/j.patrec.2018.04.006

[61] Loconsole, C., Cascarano, G.D., Lattarulo, A., *et al.*: 'A comparison between ANN and SVM classifiers for Parkinson's disease by using a model-free computer-assisted handwriting analysis based on biometric signals'. 2018 Int. Joint Conf. Neural Networks (IJCNN), Rio de Janeiro, Brazil, 2018, pp. 1–8, doi: 10.1109/IJCNN.2018.8489293

[62] Goetz, C.G., Tilley, B.C., Shaftman, S.R., *et al.*: 'Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results', *Mov. Disord.*, 2008, **23**, (15), pp. 2129–2170

[63] Buongiorno, D., Trotta, G.F., Bortone, I., *et al.*: 'Assessment and rating of movement impairment in Parkinson's disease using a low-cost vision-based system', in Huang, D.S., Gromiha, M.M., Han, K., *et al.* (Eds.): '*Intelligent computing methodologies*' (Springer International Publishing, Cham, 2018), pp. 777–788, doi: 10.1007/978-3-319-95957-3_82

[64] Bortone, I., Buongiorno, D., Lelli, G., *et al.*: 'Gait analysis and Parkinson's disease: recent trends on main applications in health care', in Masia, L., Micera, S., Akay, M., *et al.* (Eds.): '*Converging clinical and engineering research on neurorehabilitation III*' (Springer International Publishing, Cham, 2019), pp. 1121–1125, doi: 10.1007/978-3-030-01845-0_224

[65] Cascarano, G.D., Brunetti, A., Buongiorno, D., *et al.*: 'A multi-modal tool suite for Parkinson's evaluation and grading'. Proc. 2018 Italian Workshop Neural Networks, 2018, accepted publication

[66] Taylor Tavares, A.L., Jefferis, G.S., Koop, M., *et al.*: 'Quantitative measurements of alternating finger tapping in Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation', *Mov. Disord.*, 2005, **20**, (10), pp. 1286–1298

[67] Gunzler, S.A., Pavel, M., Koudelka, C., *et al.*: 'Foot-tapping rate as an objective outcome measure for Parkinson's disease clinical trials', *Clin. Neuropharmacol.*, 2009, **32**, (2), pp. 97–102

[68] Hashem, I.A.T., Yaqoob, I., Anuar, N.B., *et al.*: 'The rise of 'big data' on cloud computing: review and open research issues', *Inf. Syst.*, 2015, **47**, pp. 98–115, doi: https://doi.org/10.1016/j.is.2014.07.006

[69] Kim, H.C., Pang, S., Je, H.M., *et al.*: 'Support vector machine ensemble with bagging', in Lee, S.W., Verri, A. (Eds.): '*Pattern recognition with support vector machines*' (Springer, Berlin, Heidelberg, 2002), pp. 397–408

[70] Lei, Z., Yang, Y., Wu, Z.: 'Ensemble of support vector machine for text-independent speaker recognition', *Int. J. Comput. Sci. Netw. Secur.*, 2006, **6**, (5), pp. 163–167

[71] Ghemawat, S., Gobioff, H., Leung, S.T.: '*The Google file system*', vol. **37**, (ACM, 2003)

[72] Alham, N.K., Li, M., Liu, Y., *et al.*: 'A MapReduce-based distributed SVM ensemble for scalable image classification and annotation', *Comput. Math. Appl.*, 2013, **66**, (10), pp. 1920–1934. iCNC-FSKD 2012, doi: 10.1016/j.camwa.2013.07.015