A Comprehensive Scheme for the Objective Upper Body Assessments of Subjects with Cerebellar Ataxia

Ha Tran1, Khoa D Nguyen1, Pubudu N Pathirana1, Malcolm K Home2, Laura Power3 and David J Szmulewicz2,3,4

Abstract

Background: Cerebellar ataxia (CA) is a complex motor disorder that exhibits various symptoms such as lack of movement accuracy, delayed motion and ataxic movements associated with gait, extremity and eye.

Accurate assessment of ataxic movements forms an integral part, not only in the process of diagnosis, but also to monitor the severity of the neurodegenerative progression, particularly in a rehabilitation context. However, the current assessment schemes are mostly based on the subjective observation of experienced clinicians.

Capturing the movement during standard upper limb tests using readily available motion sensors, this paper is intended to amalgamate the sensory information to obtain a more accurate and objective form of assessment.

Methods: An assessment scheme involving an inertial measurement system and a Kinect® system was considered to quantify the degree of ataxia in four instrumented version of upper extremities tests, i.e. Finger Chase (FCT), Finger Tapping (FTT), Finger to Nose (FNT) and Dysdiadochokinesia (DDKT). Kinematic features from these tests were extracted to quantitatively define ataxic signs such as dysmetria, delay in timing, irregularity and instability. Using Feed backward feature elimination (FBE) and Quadratic discrimination analysis (QDA) and Ridge regression (RR), the features were selectively combined to improve the diagnosis and verify the association with clinical assessments by means of Leave-One-Out cross validation. Clinical ratings of the disease status were recorded using the Scale for the Assessment and Rating of Ataxia (SARA).

Results: We report statistical significance in identifying ataxia from movement features of the four tests. The combined information from the features provided a high accuracy in diagnosing CA subjects (96.4%) in addition to a promising result in predicting the severity of ataxia due to CA ($\rho = 0.8, p < 0.001$). The severity estimation was also considered in a 4-level scale to provide a rating that is familiar to the current clinically-used rating of upper limb impairments. The combination of FCT and FTT achieve the most acceptable outcome among the considered subsets of the 4 tests.

Conclusion: The analysis of ataxia can be decomposed primarily into four affected dimensions, i.e. stability, timing, accuracy and rhythmicity. In the context of upper limb tests, the results of accurate classification and prediction of severity attributed mostly to the timing. Furthermore, the underlying approach uncovers the appropriate combination with a reduced number of tests for the assessment of CA utilising the clinical resources more effectively.

Trial registration: Human Research and Ethics Committee, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia (HREC Reference Number: 11/994H/16).

Keywords: Cerebellar ataxia; Finger chase; Finger tapping; Finger to nose; Diadochokinesia; Objective assessment; Feed backward feature elimination
standard motor tasks described by Holmes [5, 6] and others almost a century ago. Deficits in the movement are described in Greek terms such as dysmetria, dyssynergia and dysrhythmia. As recognised by pioneering clinicians such as Holmes [5, 6], ataxic movements could not easily be reduced to Newtonian terms but fundamentally manifest disturbances in accuracy, timing, rhythmicity and stability of the proximal motor platform.

Upper limb ataxia is generally assessed by bedside tests such as finger chasing (FCT), finger tapping (FTT), finger to nose (FNT) and tests of fast alternating hand movements looking for dyshiychodokinesia (DDKT) [7, 8] referred to from here as "tests". The clinical evaluation of these tests is codified by means of scales such as the Scale for the Assessment and Rating of Ataxia (SARA) [7]. The SARA identifies specific aspects of motor dysfunction that should be considered when scoring the disability e.g., the SARA stipulates the overshoot/undershoot distance between subject’s finger and clinician’s finger in the finger chase test. However, there is inevitable variation in the rating of functional deficits using such clinical scales, because inherently these depend on human interpretation and subjective decisions. The SARA and conventional teaching also recommend administering several tests to characterise upper limb ataxia. While this may provide some clinical security that the deficit is revealed in several different tests it is also likely that there is redundant information.

Quantitative approaches facilitated by the recent technological advancements provide means to overcome subjectivity. Several sensing and information extracting systems have been proposed for the quantitative assessment of upper limb ataxia. For example, a push-button system capturing successful taps within a certain amount of time was considered for the FTT [9-11] so as to evaluate the time variation characteristics of ataxic movements (impaired timing). However, neither kinematic variability or stability of the platform provided by proximal joints were not considered. Inertial measurement units (IMUs), therefore, appeared as a promising means of capturing movement kinematics in multiple signal domains [12] without the requirement of sophisticated and often complex infrastructure. IMU based systems were also used effectively in FNT [13, 14] and DDKT [13] where the axes of movement characterising ataxia were captured, monitored and evaluated. The movement of the finger performing the FCT has been tracked using optoelectronic devices ranging from video cameras [15] to VICON [16] and recently Kinect© in our previous study [4] to assess delay in initiating movement and accuracy in reaching the target. While this test identifies deficits in accuracy and timing, neither ability to maintain a rhythm or the stability of the execution platform of the moving distal limb are assessed [5]. Thus it has been possible to emulate individual bedside tests through objective assessments but none of these tests appear to fully assess all aspects of upper limb ataxia (timing accuracy, rhythmicity or proximal stability). Even in those tests that addressed similar aspects, the extent to which they measure the same aspect similarly (i.e. are redundant) is not clear.

In this work, we propose an Instrumented CA Upper Limb (ICUL) System that can provide an objective assessment of ataxia and identify a combination of tests that provide sufficient information of the disability with minimal redundancy. Information obtained from sensors worn by subjects while performing the four well accepted tests of upper limb ataxia, namely FCT, FTT, FNT and DDKT to obtain instrumented scores of ataxia. The scores were modelled and tested against clinical ratings (SARA) performed at the same time. The instrumented data obtained during each ICUL test was analysed further to extract objective measures based on kinematics, entropy and other manipulations to build a feature set related to each ICUL. These features were used to find combinations of tests that gave the best discrimination between controls and ataxics, the best assessment of severity and to remove redundant information. Based on Holmes’ definitions, ataxic features were decomposed into stability, timing, accuracy and rhythmicity (STAR) dimensions. The extracted measurements from the system were grouped as per their identified STAR dimension to quantify the heterogeneous aspects intrinsic to ataxia. Feed backward feature elimination (FBE) reduced the feature space to a subset that improved the accuracy in distinguishing normal and ataxic movement. Further, for the purpose of clinical uses, the severity level of the condition was estimated using regression analyses and quantified into a 4-level scale resembling the upper limb scale of SARA.

| Table 1 Participant demographics |
|----------------------------------|
|                                | Controls | Subj with CA |
| Total subjects (M/F)            | 14(5/9)  | 41(24/20)    |
| Dominant hand, (L/R)            | 2/12     | 2/39         |
| Age, mean ± SD (years)          | 55 ± 18  | 64 ± 15      |
| SARA score, mean± SD            | -        | 14.07 ± 9.97 |
| Total score                     | -        | 3.58 ± 2.62  |
| Diagnosis:                      |          | 8/5          |
| CABV/CANVAS                     | -        | 4/10/14      |
| FA/SCAs/Others                  | -        |              |

Tran et al.

Page 2 of 14
Figure 1 Instrumented version of the upper limb assessments and the movement waveform of a control and a patient diagnosed with CA. (A) Finger Chase (ballistic, FCT) using Kinect© system. (B) An IMU sensor with tri-axial accelerometer directions \((Ax,Ay,Az)\) with a gyroscope directions \((Gx,Gy,Gz)\). (B)II. Sensor placement around the wrist. (B)III. Sensor placement around the palm. Testing with the IMU system denoting the direction of the primary movement; movement along the direction of effective axis in order to accomplish the task objectives: (C) Finger Tapping (FTT), (D) Finger to Nose (FNT), (E) Dysdiadochokinesia (DDKT).

**Methods**

**Participants**

Fourteen control subjects (mean age, 42; range, 25-68 years) and 41 subjects with CA (mean age, 59; range, 28-78 years) participated in this study (Table 1). All subjects with CA were previously diagnosed with a progressive neurodegenerative ataxia (with genotyping or other confirmatory investigations when relevant, see Table 1 - Diagnosis). The severity of ataxia was assessed using SARA [7], on the same day that objective measurements were made. The same clinician, experienced in assessing ataxia, provided all SARA scores to avoid the inevitable scoring variation that occurs when different subjects are assessed by different clinicians. SARA assessment is comprised of eight components including three that score the lower limbs (No.1-2-8), three that score the upper limbs (No.5-6-7), on score for sitting (No.3) and one score speech (No.4). The SARA scores three of the four tests of upper limb function used in this study: FCT (No.5), FNT (No.6) and DDKT (No.7). Each component receives a score in 5 levels (0 to 4 points) according to the clinician’s judgement of severity of ataxia in performing the specific test. Hence the upper limb SARA score (SARA-UL) is from 0 to 15 points. The SARA total score (SARA-Total) is calculated by the summation of the eight-component scores resulting in a maximum of 40 points. In this study, the SARA-Total (14.23 ± 9.84) and the SARA-UL (3.58 ± 2.62) were correlated with the objective assessment of severity of ataxia. This study was approved by the Human Research and Ethics Committee, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia (HREC Reference Number: 11/994H/16). Written consent was obtained from all participants.

**Assessment Protocols & Apparatus**

The ICUL consisted of four tasks performed with instrumented devices: FCT, FTT, FNT and DDKT. The FCT used a depth sensing camera (Kinect©) to capture the movements of the subject’s finger (while reaching a target on the screen) while in the other three tests, kinematic information was acquired from a 3-dimensional (3D) IMU system, BioKinTM [17] system. The test descriptions and protocols are summarised in Table 2. All the tests were performed under the supervision of an expert clinician (LP). During the test, the clinician wirelessly started and stopped recording and applied markers into the data stream, denoting specific points during the performance through a mobile application. At the end of each trial, the sensor data were uploaded to a cloud-based storage and computing platform for further analysis.

**Manifestations of Ataxia**

Following Holmes [5], we describe four domains of ataxia (using the acronym STAR). The purpose is to develop a system of assessing ataxia that reflect the following generic domains of manifestations:
Table 2 Description of the ICUL tests.

| Test                  | Device       | Setup                                                                 | Description                                                                 |
|-----------------------|--------------|----------------------------------------------------------------------|----------------------------------------------------------------------------|
| Finger Chase Test (FCT)| Kinect©      | A Microsoft Kinect© V2, a 23 inch screen and processing computer (Intel core i5) are installed approximately 1.5m away the subject. The Kinect© captures movements from a 14mm retro-reflective marker attached on the subject's index finger. A program randomly generates the target point 20 times on the monitor while projecting the finger movement on the screen. | The subject is required to point at and follow a target point on the screen using the index finger. As soon as the projected marker point touches the target point, the target disappears and reappears at a new position. The test is concluded after 20 iterations (Fig. 1(A)). |
| Finger Tapping Test (FTT) | IMU          | A sensor was worn on the dorsum of the hand as depicted in the Fig. 1(B)III. | The subject is required to tap on a tabletop using the index finger at a self-selected and uniform pace. Tapping is performed for approximately 15 seconds with the elbow and shoulder joints unsupported to assess the stability of the platform (shoulder and elbow) (Fig. 1(C)). |
| Finger to Nose Test (FNT)  | IMU          | A sensor was worn on the dorsum of the hand as depicted in the Fig. 1(B)III. | The subject's index finger moves repeatedly between the the clinician's finger and the subject's nose for approximately 15 seconds. The clinician's finger is held stationary at a position approximately 50cm in front of the subject (Fig. 1(D)). |
| Dysdiadochokinesia Test (DDKT) | IMU         | A sensor was worn on the wrist as depicted in the Fig. 1(B)III. | The subject alternates between placing one hand palm-up and palm-down on the other hand as fast and precisely as possible for approximately 10 cycles (Fig. 1(E)). |

- **Stability (S):** Lack of stability in the platform during the execution of the task (the oscillations of the movement that is not preferred).
- **Timing (T):** Error between the goal/time objective against what is achieved in a temporal context. This is likely to be affected by:
  - The time for the subject to initiate a movement.
  - The time to complete a movement/speed.
- **Accuracy (A):** Error between the goal/space objective against what is achieved in a spatial context.
- **Rhythmicity (R):** The regularity in repeated movement.

Data Preprocessing

Accelerations and angular velocities from the IMU sensor were sampled at 50Hz in the three orthogonal X, Y and Z axes. These signals were filtered by a 2nd order band-pass Butterworth filter with the cut-off frequency from 0.3Hz to 20Hz where the base band frequencies were excluded to minimise drift effects and the high frequencies were restricted to the bandwidth of human movements [18]. In the Kinect© system, the location of each randomly generated instantaneous position change was stored as a pair of position coordinates. The target position remained constant between each change in the target location, while the marker position changed. The Kinect© captured the marker position at a sampling rate of 30Hz, which was well above the sampling frequency of 20 Hz required for human movement [19].

Data analysis

Relevant objective measures extracted from each ICUL test were described and associated with the corresponding STAR classification in Table 3. For notational simplicity, the feature names are denoted as: $(FeatureName)^{Axis(L/R)}_{Test}$ with L/R indicating performance by the left or right hand.

**Finger Chase Test (FCT)**

When assessing the FCT, the clinician subjectively estimates the extent of under/overshooting in the subject’s movements relative to the moving target [7]. The ICUL system automates the assessment of FCT by considering the space-time trajectory of the marker and target. The overshoot/undershoot information of the subject movement was measured by the Dynamic time warping (DTW)-based error. The DTW was used to find the shortest path between the marker $S_m$ and target $S_r$ trajectories via their distance matrix $DS$ using dynamic programming

$$DS(i, j) = \text{dist}(S_m(i), S_r(j)) + \min\{DS(i - 1, j), \quad DS(i - 1, j - 1), \quad DS(i, j - 1)\}. \quad (1)$$

The error $DTWErr$ is calculated by summing the value of the shortest path $P$ obtained by going from the last (n; n) to the first (1; 1) element of $DS$ via adjacent elements with the smallest values. The time from establishing a new target position to the subject's initiation of movement was defined as reaction time (ReTi). This feature was obtained by cross correlating
the two time sequences representing the marker and the target movement.

\[
ReTi = \arg \max \left( \sum_{t=-\infty}^{\infty} S_m[i]S_t[i+j] \right).
\]  

(2)

The kinematic delay was obtained from the index of performance measurement described by Fitts’ law [20]. The feature is intended to capture the performance of the subject in reaching a target position outlined by \( KiDe = ID/MT \), where \( ID = \log_2(\text{distance}) \) is the index of difficulty of the task while \( di \) is the distance between the current and previous position of the target, \( ra \) is the radius of the target circle and \( MT \) is the execution time of the task by the subject. The acceleration alteration (\( \text{AcAlt} \)) computed the number of times the subjects changed their acceleration while reaching the target. The feature measures the efficiency of force applied to performing the task.

**Finger Tapping Test (FTT)**

Temporal variability is greater when ataxic subjects tap repetitively than when controls do [21]. This can be observed in the inter-tap interval (ITI) and "movement variability". The ITI is defined as the duration between successive contacts with the table and its coefficient of variation (\( CITT \)). This quantifies the variability of the tapping rhythm with respect to the tapping rate [21, 22]. Movement variability is quantified using fuzzy entropy (\( FuEn \)), obtained for each movement time series (accelerations and angular velocities). Instead of using Heaviside function, the similarity degree \( D_{pq}^m \) between two states \((p \text{ and } q)\) in the phase space is quantified using a fuzzy function \( \mu_D(d_{pq}, r) = \exp(-(d_{pq}/r)^2) \) of order 2 and tolerance \( r \). Generally, entropy allows variability to be quantified by calculating the reduction of information when the embedding dimension \( m \) increases by one [12, 23].

\[
FuEn = \ln \phi^m(r) - \ln \phi^{m+1}(r)
\]  

(3)
where,
\[
\hat{\phi}^m(n, r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left[ \frac{1}{N-m-1} \sum_{p=1, p \neq q}^{N-m} D^m_{pq} \right].
\]

(4)

Finger to Nose Test (FNT) and Tests of Dysdiadochokinesia (DDKT)
There are movement characteristics in FNT and DDKT that can be considered together in analysis. They are both repetitive movements that require a stable platform (shoulder in both cases). In both tasks, ataxia is not only characterised by variability in rhythm but also by prolonged task duration resulting from errors in movement displacement. Such characteristics are amenable to investigation using frequency domain techniques. Measurements from the accelerometer and gyroscope were analysed in terms of the resonant frequency (RF) and its magnitude (MR) using Fast Fourier Transforms (FFT) with appropriate filtering parameters (6th order bandpass Butterworth filter with the cut-off frequency region of 2-5Hz). In the FNT test, the angular accelerations and linear accelerations can be used to characterise ataxia [13] more effectively. The RF and MR of angular accelerations in the three axes were calculated as well as the linear acceleration in the X axis. Only RF was applied to the linear acceleration in the Y and Z axes. In the case of the DDKT, RF and MR of angular accelerations and linear accelerations accelerations in all axes best distinguished between ataxics and controls [13].

Statistical inferences
Statistical significance between CA and controls was identified using hypothesis tests. Normality in the variables was tested using the Shapiro-Wilk test. The t-Student’s test was applied for normally distributed variables. The Wilcoxon rank-sum test or Mann-Whitney U test was applied for variables that were not normally distributed. To test for validity, Spearman correlation was used to measure the relationship between objective measurements and clinical scales. The sample size used in this study was determined to detect the minimum effect size of 0.88 with 80% statistical power and significance level (α) of 5%. Similarly, for testing correlation, the effect size (r) of 0.30 was used. The power analysis was performed using G*Power version 3.1.9.4 [24].

Feature Selection
The four tests of upper limb ataxia produced many features. As the feature space was unlikely to be uniformly populated, there was a risk of overfitting a learning model. To overcome this and improve the prediction power the number of features were reduced using feature selection (FS) techniques. The Feed Backward Feature Elimination (FBE) [25] was employed along with three other widely-used methods involving Random Forest [26], RELIEF [27] and LASSO [28]. The central idea of FBE is to find a subset of features that increases the model’s performance. In each iteration of the process, 90% of the data was randomly selected and a feature elimination decision was made using a threshold α (significance level) on the p-value of the feature (with the null hypothesis H0 that the examining feature is independent of the predicted score given the set of currently selected features). Therefore, only features that have significant impacts on the output (p-value<α) are selected for the feature subset.

In our experiment, we repeated the process 100 times to obtain the selection frequency of each feature in estimating its significance in the assessment/diagnosis problem. Details of the process was explained in the flowchart in Fig. 2(A). Since the process is time consuming, computational performance was improved by employing the Parallel Computing Toolbox of MATLAB version 2019b to simultaneously execute the computations.

Discrimination & Severity Analysis
Features with high selection frequency from the FBE were used to classify control and CA groups and predicting the severity of ataxia. Classification models for diagnosis included Linear Discrimination (LD) [29], Quadratic Discrimination Analysis (QDA) [30], Support Vector Machine (SVM) [31], K-Nearest Neighbour (KNN) [32], Leave-One-Out (LOO) cross validation estimated the generalised performance of the diagnostic model as well as the severity predicting regression model [33]. The effectiveness of the model was evaluated through a number of statistical measurements including accuracy (ACC), F1-score, the stability of the model by the area under the Receiver Operating Characteristics curve (AUC), the sensitivity measure (or Recall), and Precision.

For regression analysis, we employed the Ridge regression method to correlate the proposed features with the SARA scores. This model avoids over-fitting working with small data sets by forming a linear model to estimate the severity for the given input feature vector. In order to generate a general instrumented score, a severity scale that mirrored the SARA upper limb scores was developed. As there is no ataxic subjects in the cohort were rated with a SARA score of 4 (i.e all subjects ranged from 0-3), the instrumented severity scale was limited to 4 levels defined as follows:

- **Level 0**: Normal, no dysesthetia, tremor or irregularities.
- Level 1: Mild, light dysmetria in small range, low amplitude tremor or slight irregular motion.
- Level 2: Moderate, clear dysmetria, tremor or clearly irregular motion.
- Level 3: Severe, dysmetria in large range, high amplitude tremor or very irregular motion.

**Results**

**Feature Significance**

Table 4 shows the 31 (out of 62) objective features generated during the performance of FCT, FTT, FNT and DDKT that reached statistical significance \(p < 0.05\). These features represent movement characteristics of ataxic subjects that differ significantly from controls. Movements performed by subjects with CA were significantly slower (e.g. KiDe, \(p < 0.001\)) with longer reaction times (e.g. ReTi, \(p < 0.001\)) suggesting that a longer time is required to recognize the new target and react. The movements of controls were relatively more complex movements (entropy measures, \(p < 0.05\)) than movements of ataxic subjects and had less functional variability (e.g. DTWEr, \(p < 0.001\)). There were mea-

| Test | Feature name | Subjects with CA | Controls | ES & p-value | CC & p-value | CC & p-value |
|------|--------------|------------------|----------|--------------|--------------|--------------|
| FCT  | KiDeL \(L\)  | 2.683±0.511      | 3.494±0.431 | 1.646 (<0.001) | -0.572 (0.003**) | -0.663 (<0.001) |
|  | KiDeR \(R\)  | 2.517±0.553      | 3.575±0.352 | 2.070 (<0.001) | -0.504 (0.010*)  | -0.380 (0.061) |
|  | ReTiX \(X\)  | 1.074±0.292 \((x10^3)\) | 0.761±0.083 \((x10^3)\) | 1.217 (<0.001) | 0.610 (0.001**) | 0.659 (<0.001) |
|  | ReTiY \(Y\)  | 1.166±0.332 \((x10^3)\) | 0.740±0.084 \((x10^3)\) | 1.462 (<0.001) | 0.358 (0.079)  | 0.223 (0.284) |
|  | ReTy \(R\)  | 1.076±0.294 \((x10^3)\) | 0.762±0.085 \((x10^3)\) | 1.211 (<0.001) | 0.595 (0.002**) | 0.641 (<0.001) |
|  | ReTy \(R\)  | 1.170±0.334 \((x10^3)\) | 0.743±0.083 \((x10^3)\) | 1.456 (<0.001) | 0.360 (0.077)  | 0.228 (0.272) |
| TTT  | DTWEr \(L\)  | 1.978±0.648 \((x10^4)\) | 1.237±0.182 \((x10^4)\) | 1.299 (<0.001) | 0.433 (0.031*)  | 0.394 (0.051) |
|  | DTWEr \(X\)  | 2.362±0.853 \((x10^4)\) | 1.201±0.291 \((x10^4)\) | 1.539 (<0.001) | 0.305 (0.138)  | 0.086 (0.682) |
|  | DTWEr \(Y\)  | 2.277±0.924 \((x10^4)\) | 1.301±0.253 \((x10^4)\) | 1.202 (<0.001) | 0.684 (<0.001) | 0.552 (0.004**) |
|  | DTWEr \(Z\)  | 2.705±1.117 \((x10^4)\) | 1.471±0.250 \((x10^4)\) | 1.209 (<0.001) | 0.290 (0.159)  | 0.109 (0.604) |
|  | DDTWPEr \(L\) | 0.335±0.125 \((x10^3)\) | 0.221±0.047 \((x10^3)\) | 0.984 (<0.001) | 0.522 (0.008**) | 0.507 (0.010*) |
|  | DDTWPEr \(X\) | 0.357±0.145 \((x10^3)\) | 0.209±0.029 \((x10^3)\) | 1.163 (<0.001) | 0.553 (0.004**) | 0.271 (0.191) |
|  | DDTWPEr \(Y\) | 0.257±0.117 \((x10^3)\) | 0.161±0.036 \((x10^3)\) | 0.926 (<0.001) | 0.275 (0.184)  | 0.317 (0.122) |
|  | DDTWPEr \(Z\) | 0.267±0.120 \((x10^3)\) | 0.141±0.025 \((x10^3)\) | 1.191 (<0.001) | 0.202 (0.332)  | 0.105 (0.618) |

Data are shown in mean ± standard deviation.

Abbreviations. CA Subjects with cerebellar ataxia, HC Controls, ES Effect size, CC Correlation coefficient (Spearman).

\*p-value < 0.05, **p-value < 0.01.
ures in the non-primary axis whose values in controls differed significantly from ataxic subjects. These differences most likely arose from instability in proximal or stabilising joints of individuals with CA. The clinical validity of these measures was assessed by correlating with the SARA scores (SARA-Total and SARA-UL; see last two columns of Table 4). Some FCT movement characteristics were moderately ($p < 0.01$) to significantly ($p < 0.001$) correlated with the SARA ratings. Entropy features of FTT correlated moderately with SARA, while DDKT and FNT features correlated weakly with SARA ($p < 0.05$).

Selection Frequency of Features

Figure 2(E) shows the selection frequency following 100 iterations of the FBE process applied to 22 features with the highest contribution (first 22 features) in the combined test model. The frequency implies the contribution of each feature in estimating severity of CA. Higher selection frequency, implies a greater possibility of the feature selected in the final subset. Of note, FCT provided more important features (including ReTi, the feature with the highest selection frequency) than other tests and all FCT movement features appear in the chart. As discussed later this reflects the importance of FCT in the objective ataxia score. In comparison, FNT contributed the least to the selected feature subset. Despite fewer features (3), FTT features were selected with higher frequency than DDKT or FNT related features.

Figure 3 plots the classification performance (Y axis) of the four feature selection methods against the number of selected features (X axis). Here, the aim was to find the smallest subset of features that produced a high performance (accuracy) in diagnosing CA. FBE outperformed LASSO, RELIEF and Random Forest,
CA Diagnosis and SARA based Severity Estimation

The accuracy of the system in making a binary diagnostic classification (into CA subjects and controls) can be considered with Precision and Recall values. Precision is measured by expressing the number of CA subjects correctly identified as a fraction of the total of subjects identified as CA. Recall expresses the number of CA subjects correctly identified as fraction of the total of actual CA subjects. Therefore, a model with Precision and Recall values closer to 1 is considered an effective model for diagnosis. Furthermore, due to the imbalanced number of CA subjects and controls, Matthews correlation coefficient (MCC) was used to assess the effect of the higher number of ataxic subjects (CA) on the model's accuracy. The value range of MCC is from -1 to 1, where 1 depicts a perfect prediction. The diagnostic performance of four learning models (QDA, LD, SVM, KNN) and four feature selection methods (FBE, LASSO, RELIEF, RF) were compared providing a 96.4% accuracy with the first 22 of the 64 (top 34%) features with the highest selection frequency. The accuracy was low (ACC 83%) with the first 5 features. As the number of features increased, the performance of FBE fluctuated around 95.4% (std. ± 1.4%) with performance comparable to other methods. List of selected features are shown in the bar chart of Fig 2(E).

The predicted scores, peak moments of tests of assessment are in investigated in the last subsection.

| Classifier | FS | Recall | Precision | MCC | ACC | AUC |
|------------|----|--------|-----------|-----|-----|-----|
| QDA        | FBE | 0.98   | 0.98      | 0.90| 96.4| 0.97|
|            | RF  | 0.93   | 0.98      | 0.81| 92.7| 0.96|
|            | RELIEF | 0.93 | 0.90      | 0.66| 87.3| 0.88|
|            | LASSO | 0.85 | 0.90      | 0.54| 81.8| 0.87|
| LD         | FBE | 0.90   | 0.93      | 0.67| 87.3| 0.85|
|            | RF  | 0.85   | 0.90      | 0.54| 81.8| 0.80|
|            | RELIEF | 0.88 | 0.78      | 0.19| 72.7| 0.65|
|            | LASSO | 0.83 | 0.79      | 0.20| 70.9| 0.62|
| SVM        | FBE | 0.83   | 0.92      | 0.57| 81.8| 0.85|
|            | RF  | 0.88   | 0.92      | 0.64| 85.5| 0.89|
|            | RELIEF | 0.90 | 0.97      | 0.78| 90.9| 0.95|
|            | LASSO | 0.90 | 0.93      | 0.67| 87.3| 0.90|
| KNN        | FBE | 0.95   | 0.98      | 0.86| 94.5| 0.97|
|            | RF  | 0.90   | 0.93      | 0.67| 87.3| 0.90|
|            | RELIEF | 0.88 | 0.88      | 0.52| 81.8| 0.93|
|            | LASSO | 0.95 | 0.95      | 0.81| 92.7| 0.94|

The QDA+FBE pair outperformed the others in diagnostic performance with a greater accuracy (ACC 96.4%, Recall 0.98, Precision 0.98) and reliability (AUC 0.97 and MCC 0.90). This classification can be visualised by plotting the first 3 principle components of a Principal Component Analysis (PCA) (Fig. 4(A)). In summary, CA can be identified (diagnosed) with a high degree of accuracy (ACC > 92%, Precision & Recall > 0.9, MCC > 0.7) using several models (QDA+FBE, QDA+RF, KNN+LASSO, KNN+FBE) generated from extracted features. The QDA provided a flexible decision boundary for assessing the influence of each clinical test on the capacity to accurately separate CA subjects from controls. Table 6 shows that FCT performed better than the others (ACC 92.7%, Recall 0.95, Precision 0.95) despite the scarcity of rhythmicity features in the selection (Table 4). Notwithstanding the performance of FCT, all tests combined provided the most accurate model in diagnosing CA subjects (ACC 96.4% compared to ACC 92.7%, Table 6) and was also less affected by the imbalance between individuals with CA and controls (MCC 0.90 compared to MCC 0.81, Table 6). This superior performance of the combined tests over FCT highlights the importance of having all domains presented: rhythmicity features which were not in FCT were provided by other tests. However, domains can be provided by more than one test; for an instance, rhythmicity is present in FFT, FNT & DDKT and stability is provided by all 4 tests. This raises the question of the need for all tests to accurately assess the severity of ataxia. Thus, we map features and tests to the STAR dimensions and then different combinations of tests are investigated in the last subsection.

As previously shown, the values predicted by the model were highly correlated with the SARA score. The predicted scores, ps (all tests, Table 9), of the subjects were plotted against their corresponding SARA scores averaged across the upper limb tests (mean SARA_UL) in Fig. 5(A). The boxplots represent the
also be related to the presence and severity of ataxia. Their contribution to the feature selection process can be considered as follows: $ps < 4$ belongs to the normal group (level 0), $4 \leq ps < 7$ belong to the mild group (level 1), $7 \leq ps < 10$ belongs to the moderate group (level 2) and $ps \geq 10$ belongs to the severe group (level 3). The agreement matrix in Fig. 5(B) outlines the mapping of the predicted scores into each clinical severity level. In particular, normal subjects can be predicted with a high degree of accuracy (90%) from the underlying system. Two subjects who were diagnosed as normal were detected having mild ataxia due to CA (10%). No subject that was diagnosed as normal to mild condition (level 1) was classified to moderate (level 2) nor to severe group (level 3) and no CA patient was classified as normal by the proposed system.

Disability Association to The STAR Dimensions

The extracted features were assigned to one of the proposed Holmesian dimensions (STAR) of ataxia. Detail of this clustering is presented in Table 3. Selected features in each dimension of the STAR together with their contribution to the feature selection process can also be related to the presence and severity of ataxia.

Using this approach, it is possible to attribute the contribution of each STAR dimension to the overall diagnosis of ataxia (Fig. 2(A)), with the stability features contributing most (41% compared to 25% from Rhythmicty, 20% Timing and 14% Accuracy). Most of stability features were from the DDKT and FNT.

Considering dimensional aspect of the extracted features, Table 8 denotes the correlation between the STAR features and the three SARA scores, i.e. the SARA-Total and the SARA-UL (in terms of sum and mean). This revealed that the features corresponding to timing led to the highest correlation levels with the SARA scores considered (0.77, 0.87 and 0.85). Rhythmicty dimension had the lowest correlations (0.35, 0.47 and 0.38) of the four dimensions (Table 8).

Combination of Tests

In order to determine whether SARA scores can be predicted with fewer clinical tests, the performance of different test combinations was investigated. As discussed in the STAR analysis, only the FCT provided features that corresponded to the accuracy dimension. The presence of FCT features, therefore, is essential to ensure all dimensions of ataxia are addressed in the given form of instrumentation. The test groupings are considered as follows:

- **Group 1 (G1):** FCT & FTT
- **Group 2 (G2):** FCT & FNT
- **Group 3 (G3):** FCT & DDKT
- **Group 4 (G4):** FCT & FNT & DDKT
- **Group 5 (G5):** FCT & FTT & DDKT
- **Group 6 (G6):** FCT & FTT & FNT

Performance of different groupings is depicted in Table 9. As the result, the highest accuracy in the diagnosis of CA was observed in G1 (FCT and FTT). Other performance parameters such as AUC, sensitivity, and precision indicated that G1 yielded higher quality of predictions and better reproducibility as compared to other combinations (group 2 to 6). Clear separation between subjects with CA and controls is evident from Fig. 4(B). Additionally, the classification achieved by G1 was comparable with that of the combination of

| Table 7 | Common test-based features selected from the 4 FS methods. |
|---------|-----------------------------------------------------|
|         | FCT | FTT | FNT | DDKT |
| S       | n/a | n/a | $R_{FA}^{AAX}$ | $R_{DDK}^{AAX}$ |
|         |     |     | $R_{FA}^{AY}$ | $R_{DDK}^{AAX}$ |
| T       | $K_{BTC}^{XY}$ | $CIT_{FTT}$ | $R_{FA}^{AY}$ | $R_{DDK}^{AAX}$ |
| A       | n/a | n/a | n/a | n/a |
| R       | n/a | $Fu_{En}^{CyX}$ | n/a | n/a |

n/a: not available

| Table 8 | Regression analysis of STAR features and SARA. |
|---------|------------------------------------------------|
| Statistical Measure | SARA-Total | SARA-UL | SARA-UL |
| | (sum) | (mean) | |
| S | Agreement (%) | - | - | 57% |
|   | R-squared | 0.52 | 0.61 | 0.62 |
|   | Corr. coef. | 0.69 | 0.69 | 0.69 |
| T | Agreement (%) | - | - | 75% |
|   | R-squared | 0.64 | 0.82 | 0.84 |
|   | Corr. coef. | 0.77 | 0.87 | 0.85 |
| A | Agreement (%) | - | - | 56% |
|   | R-squared | 0.48 | 0.54 | 0.61 |
|   | Corr. coef. | 0.55 | 0.55 | 0.55 |
| R | Agreement (%) | - | - | 40% |
|   | R-squared | 0.33 | 0.43 | 0.39 |
|   | Corr. coef. | 0.35 | 0.47 | 0.38 |

| Table 9 | Performance of different combination of tests. |
|---------|------------------------------------------------|
| Group | ACC | AUC | Recall | Precision | F1-score | CC |
| G1 | 96.4% | 0.97 | 0.95 | 1.00 | 0.97 | 0.80** |
| G2 | 85.5% | 0.95 | 0.95 | 0.99 | 0.92 | 0.40 |
| G3 | 86.4% | 0.70 | 0.98 | 0.78 | 0.68 | 0.36 |
| G4 | 87.3% | 0.89 | 0.95 | 0.82 | 0.88 | 0.48 |
| G5 | 92.0% | 0.90 | 0.98 | 0.93 | 0.95 | 0.53 |
| G6 | 92.7% | 0.92 | 0.95 | 0.95 | 0.95 | 0.60* |
| All tests | 96.4% | 0.96 | 0.98 | 0.98 | 0.98 | 0.68** |
| CC | Correlation coefficient (Spearman), *p-value < 0.05, **p-value < 0.001. |
all tests, but with a superior effect size of the correlation. The STAR rating of G1 is depicted in Fig. 2(C)) as 32% timing, 31% stability, 22% accuracy and 15% rhythmicity.

In considering diagnostic accuracy, as both sensitivity and accuracy are important, we can consider the F1-score as the performance index by considering the relationship between the precision and specificity. Therefore, in order to evaluate the performance of the models with the same accuracy (all tests and G1), the F1-score, which is the harmonic mean of the sensitivity and precision [34] was obtained. The higher the value of F1-score infers a superior model. In that context, the performance in terms of diagnosis using all tests is marginally better than that is for G1 (F1-score = 0.98 compared to 0.97). However, considering the correlation between the predicted scores from the regression model, G1 yielded the highest correlation with the SARA-Total (Table 9, coefficient of 0.8 compared to 0.68 when using all tests). Therefore, in the instrumented system, the FCT and FTT combination significantly agreed with the clinical assessment of ataxia in the upper limb.

Discussion
Ataxia has proved hard to define, but to misquote Justice Stewart [35], most expert neurologists “know it when they see it”. The standard bedside tests develop over 100 years ago bring out the essence of ataxia and allow it to be recognised by the examining clinician. These tests emphasise the common features referenced in STAR, and remarked on nearly 100 years ago by Holmes. However, each test emphasises different STAR domains and thus begs the questions of which are the most useful in identifying ataxia and how much redundancy is there in these tests. There is also no easy means of scoring the severity of bedside test. Clinical scales in part address this but inevitably they are subjective and descriptive without addressing the essence of ataxia. Emulation of each bedside test using objective measurement obtained from sensors would in part address these issues but would not allow each test to be directly compared to allow to assess each test contribution to the diagnosis or measuring severity of ataxia. Previous studies have already shown that each individual bedside test can be emulated using features derived from sensors worn while subjects perform the bedside test. However combining several tests using comparable scoring systems would go a substantial way to addressing all the issues raised above. Thus, the aim of this study was to obtain instrumented data from subjects performing four bedside tests (FCT, FTT, FNT and DDKT) and use features from these data to model the SARA-Total and the SARA-UL scores. Approximately half of the features were significantly correlated with the two SARA scores with the highest correlation of individual features being 0.68 with the SARA-Total and 0.66 with the SARA-UL ($p < 0.001$). The feature set was further refined to 22 features and using several different learning models it was possible to identify (diagnose) CA accurately (Table 5)) using these 22 extracted features.

Not all bedside tests contributed equally to the performance of these models. FCT contributed the most features as well as the most frequently selected features and on its own performed nearly as well as the 4 tests combined. FCT combined with FTT provided enough features to perform as well as the combined feature set. One conclusion is that FCT was necessary because it was the only test that included the accuracy domain from STAR. This may be in part self-fulfilling and reflect aspects of the STAR criteria but future studies exploring different definitions of accuracy or other tests which measure accuracy could address this issue. Even though accuracy was only present in FCT, features related to the accuracy dimension were not selected in the list of common features (Table 7). One possible explanation is that accuracy is highly correlation with timing features which may in turn have contributed to the exclusion of this dimension in LASSO. It is also noteworthy that kinematic delay in FCT contributed the most to the performance of the two models and the number of timing features significantly increased and were the highest proportion of features when FCT & FTT were combined. Further, the predicted values from the regression model that used timing features demonstrated the highest correlation with the SARA scores (Table 8) indicating the important role of timing in the clinical assessment of ataxia. Consistently selected features were obtained by extracting common features from the four FS methods. It also should be noted that features belonging to the timing domain were consistently selected from 3 out of 4 tests. They were also significant features in the combined model of all the tests and G1 model (Fig. 2(E,D)).

Another conclusion from this study is that there is redundancy in the bedside tests and not all are required to identify the presence and severity of ataxia. Multiple tests generated a plethora of features each representing aspects ataxic movements but also likely containing redundancy. The performance analysis of subsets of the tests uncovered the optimal combination of information that essentially led towards the reduction of tests. Different groupings result in feature combinations that can improve or decrease the performance of learning models (Table 9) and decreasing the number of features without affecting the performance of the learning model infers that redundant information has been removed. Combination of FCT and
FTT alone did not degrade diagnostic performance and slightly improved correlation in severity estimation in comparison to the performance of all tests combined. On the other hand, the FCT and DDKT combination was the lowest accuracy in identifying ataxia.

While the SARA prescribes that the examiner should evaluate the a) accuracy in reaching target in the FCT; b) the speed or time required to perform the DDKT; c) the amplitude of the kinetic tremor in the FNT, clinical assessment is blind in what features are found to best correlate with SARA scores. It is thus of interest that not only features that clinicians are explicitly directed to assess (e.g. accuracy in the FCT) were captured but there were also added features, e.g. initiation delay. As the instrumented test depends on these features to accurately model the SARA, this extra information is presumably identified and accounted for (possible subconsciously) by an experienced clinician even if it is not part of their explicit evaluation.

Features could be sorted into the four ataxia domains (STAR). This was most straightforward in the case of FCT, whose features could be readily placed into a STAR domain according to its physical meaning. In the case of features from FTT, FNT and DDKT, their attribution to a specific STAR domain was according to whether the feature was more related to the primary or the secondary axis of movement. The former corresponds to movements along the direction of the axis most related to accomplishing the task objectives, e.g. the upward/downward movement in tapping or the rotation of the forearm in DDKT (Fig. 1). Secondary-axis movements mostly occur because of instability of the execution platform, i.e. the proximal joints (shoulder or elbow) which must be stable for accuracy of the moving distal hand or wrist. Therefore, significant differences in secondary axes motion in ataxic and control subjects were attributed to instability in this platform. Due to the factor of repetition, the primary movement is required to adhere to a self-defined rhythm [7]. Measures pertaining to this axis can be used to infer the deficits in rhythmicity or timing. In the frequency analysis, timing aspects or "how quick is the movement performed" were described by the RF, whereas MR indicated the intensity of the rhythmic movement [13] which was considered as a measure of rhythmicity.

Learning models will always be improved with more subjects. Nevertheless, a cohort of people with CA of this size is relatively large in comparison to earlier studies of ataxia [14, 16, 36]. Furthermore, power analysis and rigorous cross validation process validated the reliability and statistical significance necessary for assertions of clinical validity. There is an assumption that “all cerebellar ataxia is the same” and it is possible, indeed likely, that the presence of somatosensory impairment, vestibular involvement or other central nervous system (CNS) lesions may affect objective assessment of ataxia. One of the motivations for producing more precise means of assessing ataxia is to establish whether the factors that might differentiate ataxia associated with other neural lesions might differ from “pure” cerebellar ataxia. This would be a subject of future studies. In a similar vein more severe ataxia reflected by SARA scores > 3 would be important in future studies. A potential direction of further research would explore the combination of FCT and FTT as a mechanism of capturing the progression of the disease in a longitudinal study. With the rapid advancement in pervasive Internet-of-Things (IoT) technologies, capturing the severity of CA subjects more regularly in their natural environment (non-clinical setting) and monitor the progress remotely will inevitably enable more personalized health care with effective rehabilitation programs.

**Conclusion**

This study demonstrates that with appropriate combinations of different instrumented functional tests, the diagnosis and assessment of ataxia can not only be facilitated with greater accuracy, but also allows the use of clinician’s time more effectively without compromising the quality of care. The instrumented assessment scheme proposed was based on the four widely-used motor tests of upper limb functionality. The system was able to support clinical decision making with fewer number of features selected from the conventional execution of these tests. The features were grouped and evaluated through the proposed definition of the ataxic manifestations (STAR) in a quantitative form which provided plausible interpretation of ataxia. In the scope of upper limb assessments, the characteristics belonging to timing resulted in the highest association with the SARA total score. A 4-level discrete form of severity rating scale was introduced to be in line with the conventional scale of SARA. This further confirmed the agreement with the current practice of clinical assessments and provided a severity estimation within acceptable levels of deviations. The other important finding of this study is that the FCT and FTT were identified as the most suitable combined assessments that presented highly accurate CA diagnosis and severity estimation among other combinations of tests. The reduction of tests would potentially lead to a more cost-effective assessment strategies to be performed in clinical practices where resources such as clinician time and the number of patient visits are often limited.
Abbreviations
CA: Cerebellar Ataxia; FCT: Finger chase test; FTT: Finger tapping test; FNT: Finger to nose test; DKDT: Tests of Dysdiadochokinesia; ICUL: Instrumented CA Upper Limb System; SARA: Scale for the Assessment and Rating of Ataxia; SARA-UL: Total score of upper limb tests in SARA; SARA-Total: Total scores of all tests in SARA; STAR: Manifestations of Ataxia (Stability, Timing, Accuracy, Rhythmicity); FBE: Feed backward feature elimination; LASSO: Least absolute shrinkage and selection operator; QDA: Quadratic discriminant analysis; RR: Ridge regression; SVM: Support Vector Machine; KNN: K-Nearest Neighbour; LD: Linear Discrimination; LOO: Leave-One-Out cross validation; IMU: Inertial measurement unit; FFT: Fast Fourier Transform; ACC: Accuracy; AUC: Area under the Receiver Operating Characteristic curve; MCC: Matthews correlation coefficient; ES: Effect size; CC: Spearman Correlation coefficient; CNS: Central nervous system.

Acknowledgements
We extend our sincere appreciation to all the patients and controls who participated in this study.

Author contributions
DJJS, MKH and PNP conceived and designed the clinical experiments; LP conducted the clinical testing; HT and KDN analysed the data; HT, KDN, PNP and MKH wrote the paper. All authors read and approved the final manuscript.

Funding
The work is conducted with the support of Florey Institute of Neuroscience and Mental Health, Melbourne, Australia. This was supported by the National Health and Medical Research Council GNT1101304 and APP1129595 and CSIRO’s Data61 while the clinical trials was undertaken in the Royal Victorian Eye and Ear Hospital (RVEEH).

Availability of data and materials
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics approval and consent to participate
Written informed consent for publication was obtained from all participants. Declaration of Helsinki prior to participation and the study was approved by the Human Research and Ethics Committee, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia (HREC Reference Number: 11/994H/16).

Consent for publication
All study participants provided consent for publication of data and images.

Competing interests
Pubudu N. Pathirana was involved in the initial design and development of BioKin2.7.10 as a data collection platform. A number of academic research outcomes have been published with Pubudu N. Pathirana as a co-author, solely outlining the novelties on various signal and data processing techniques rather than on the data collecting platform of BioKin2.7.10. Other authors declare that they have no competing interests.

Author details
1 School of Engineering, Deakin University, Warrnambool, Victoria, Australia. 2 Florey Institute of Neuroscience and Mental Health, Royal Parade, 3052 Parkville, Victoria, Australia. 3 Balance Disorders & Ataxia Service, Royal Victorian Eye and Ear Hospital (RVEEH), Gisborne St, 3002 East Melbourne, Victoria, Australia. 4 Cerebellar Ataxia Clinic, Alfred Hospital, Commercial Road, 3004 Prahran, Victoria, Australia.

References
1. Morton, S.M., Bastian, A.J.: Cerebellar control of balance and locomotion. The neuroscientist 10(3), 247–259 (2004)
2. Nguyen, N., Phan, D., Pathirana, P., Horse, M., Power, L., Szmulewicz, D.J.: Quantification of axial abnormality due to cerebellar ataxia with inertial measurements. Sensors 18(9), 2791 (2018)
3. Aclerhammer, H., Hertig, I.: The contribution of the cerebellum to speech processing. Journal of Neurolinguistics 13(2-3), 95–116 (2000)
4. Tran, H., Pathirana, P.N., Horse, M., Power, L., Szmulewicz, D.J.: Quantitative evaluation of cerebellar ataxia through automated assessment of upper limb movements. IEEE Transactions on Neural Systems and Rehabilitation Engineering 27(5), 1001–1011 (2019)
5. Holmes, G.: The croonian lectures on the clinical symptoms of cerebellar disease and their interpretation. Cerebellum 6(2), 148–53 (2007)
6. Holmes, G.: The symptoms of acute cerebellar injuries due to gunshot injuries. Brain 40(4), 461–535 (1917)
7. Schmitz-Hubsch, T., Du Montcel, S.T., Balko, L., Benciano, J., Boesch, S., Depondt, C., Gianf, P., Gobain, C., Infante, J., Kang, J.-S., et al.: Scale for the assessment and rating of ataxia: development of a new clinical scale. Neurology 66(11), 1717–1720 (2006)
8. Trouillas, P., Takayanagi, T., Hallett, M., Currier, R., Sbramony, S., Wesnes, K., Koeck, K.T., Verdun, M., Mansanet, S., Gomez, C., et al.: International cooperative ataxia rating scale for pharmacological assessment of the cerebellar syndrome. Journal of the Neurological Sciences 145(2), 205–211 (1997)
9. Shimoyama, I., Nishio, T., Uezuma, K.: The finger-tapping test: a quantitative analysis. Archives of Neurology 47(6), 681–684 (1990)
10. Notermans, N., Van Dijk, G., Van der Graaf, Y., Van Gijn, J., Wokke, J.: Measuring ataxia: quantification based on the standard neurological examination. Journal of Neurology, Neurosurgery & Psychiatry 57(1), 22–26 (1994)
11. Austin, D., McNamee, J., Klein, K., Jimison, H., Pavel, M.: A statistical characterization of the finger tapping test: modeling, estimation, and applications. IEEE Journal of Biomedical and Health Informatics 19(2), 501–507 (2014)
12. Nguyen, K.D., Pathirana, P.N., Horse, M., Power, L., Szmulewicz, D.J.: Entropy-based analysis of rhythmic tapping for the quantitative assessment of cerebellar ataxia. Biomedical Signal Processing and Control 59, 101916 (2020)
13. Kishna, R., Pathirana, P.N., Horse, M., Power, L., Szmulewicz, D.J.: Quantitative assessment of cerebellar ataxia, through automated limb functional tests. Journal of Neuromuscular and Rehabilitation 16(1), 31 (2019)
14. Martinez-Manzanares, O., Lawerence, T., Blok, H., Luning, R., Brandsma, R., Shal, D., Mauve, N.: Instrumented finger-to-nose test classification in children with ataxia or developmental coordination disorder and controls. Clinical Biomechanics 60, 51–50 (2018)
15. Bastian, A.J., Martin, T., Keating, J., Thach, W.: Cerebellar ataxia: abnormal control of interaction torques across multiple joints. Journal of Neurophysiology 76(1), 402–500 (1996)
16. Menegoni, F., Milone, E., Torti, C., Galli, M., Bigoni, M., Baudo, S., Mauro, A.: Quantitative evaluation of functional limitation of upper limb movements in subjects affected by ataxia. European Journal of Neurology 16(2), 232–239 (2009)
17. Elanayake, S.W., Morris, A.J., Forrester, M., Pathirana, P.N.: Blokin: an ambulatory platform for gait kinematic and feature assessment. Health Technol Lett. 2(1), 40–45 (2015)
18. Boute, C.V., Koekkoek, K.T., Verdun, M., Kode, R., Janssen, J.D.: A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. IEEE Trans. Biomed. Eng. 44(3), 136–147 (1997)
19. Boute, C.V., Koekkoek, K.T., Verdun, M., Kode, R., Janssen, J.D.: A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. IEEE Transactions on Biomedical Engineering 44(3), 136–147 (1997)
20. Wobbrock, J.O., Cutrell, E., Harada, S., MacKenzie, I.S.: An error model for pointing based on fits’ law. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1613–1622 (2008)
21. Spencer, R.M., Zelaznik, H.N., Diedrichsen, J., Ivy, R.B.: Disrupted timing of discontinuous but continuous movements by cerebellar lesions. Science 300(5624), 1437–1439 (2003)
22. Schief, J., Spencer, R., Zelaznik, H., Ivy, R.: Timing of rhythmic
movements in patients with cerebellar degeneration. Cerebellum 6(3), 221–231 (2007)

23. Chen, W., Wang, Z., Xie, H., Yu, W.: Characterization of surface emg signal based on fuzzy entropy. IEEE Trans. Neural Syst. Rehabil. Eng. 15(2), 266–272 (2007)

24. Faul, F., Ensefelder, E., Lang, A.-G., Buchner, A.: G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39(2), 175–191 (2007)

25. Karegowda, A.G., Manjunath, A., Jayaram, M.: Comparative study of attribute selection using gain ratio and correlation based feature selection. International Journal of Information Technology and Knowledge Management 2(2), 271–277 (2010)

26. Bi, X.-A., Hu, X., Wu, H., Wang, Y.: Multimodal data analysis of alzheimer's disease based on clustering evolutionary random forest. IEEE Journal of Biomedical and Health Informatics (2020)

27. Kira, K., Rendell, L.A.: A practical approach to feature selection. In: Machine Learning Proceedings, pp. 249–256. Elsevier, ??? (1992)

28. Theilhac, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58(1), 267–288 (1996)

29. Mitteroecker, P., Bookstein, F.: Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics. Evolutionary Biology 38(1), 100–114 (2011)

30. Sivastava, S., Gupta, M.R., Fábián, B.A.: Bayesian quadratic discriminant analysis. Journal of Machine Learning Research 8(Jun), 1277–1305 (2007)

31. Suykens, J.A., Vandewalle, J.: Least squares support vector machine classifiers. Neural processing letters 9(3), 293–300 (1999)

32. Cunningham, P., Delany, S.J.: K-nearest neighbour classifiers. Multiple Classifier Systems 34(8), 1–17 (2007)

33. Picard, R.R., Cook, R.D.: Cross-validation of regression models. Journal of the American Statistical Association 79(387), 575–583 (1984)

34. Oleh, U., Okoro, C.: Evaluating measures of indicators of diagnostic test performance: fundamental meanings and formulas. J Biom Biomath 3(1), 2 (2012)

35. Lattman, P.: The origins of justice stewart’s "i know it when i see it.". Wall Street Journal (2007)

36. Summa, S., Schiavinato, T., Bernava, G.M., Romano, A., Favetta, M., Valente, E.M., Bentini, E., Castelli, E., Petrana, M., Poggio, G., et al.: Development of Saraffome: A novel, well-accepted, technology-based assessment tool for patients with ataxia. Computer Methods and Programs in Biomedicine 188, 105257 (2020)