Instrumented classification of patients with early onset ataxia or developmental coordination disorder and healthy control children combining information from three upper limb SARA tests

Dominguez-Vega, Zeus T.; Dubber, D.; Elting, Jan Willem J.; Sival, D. A.; Maurits, Natasha M.

Published in:
European Journal of Paediatric Neurology

DOI:
10.1016/j.ejpn.2021.07.009

Publication date:
2021

European Journal of Paediatric Neurology 34 (2021) 74–83

Contents lists available at ScienceDirect

European Journal of Paediatric Neurology

Instrumented classification of patients with early onset ataxia or developmental coordination disorder and healthy control children combining information from three upper limb SARA tests

Zeus T. Dominguez-Vega*, D. Dubber, Jan Willem J. Elting, D.A. Sival, Natasha M. Maurits

University of Groningen, University Medical Center Groningen, Department of Neurology, Groningen, the Netherlands

ARTICLE INFO

Article history:
Received 3 November 2020
Received in revised form
28 June 2021
Accepted 25 July 2021

Keywords:
Early onset ataxia
Developmental coordination disorder
Inertial measurement units
Kinematic features

ABSTRACT

Background: Early Onset Ataxia (EOA) and Developmental Coordination Disorder (DCD) share several phenotypical characteristics, which can be clinically hard to distinguish.
Aim: To combine quantified movement information from three tests obtained from inertial measurements units (IMUs), to improve the classification of EOA and DCD patients and healthy controls compared to using a single test.
Methods: Using IMUs attached to the upper limbs, we collected data from EOA, DCD and healthy control children while they performed the three upper limb tests (finger to nose, finger chasing and fast alternating movements) from the Scale for the Assessment and Rating of Ataxia (SARA) test. The most relevant features for classification were extracted. A random forest classifier with 300 trees was used for classification. The area under the receiver operating curve (ROC-AUC) and precision-recall plots were used for classification performance assessment.
Results: The most relevant discerning features concerned smoothness and velocity of movements. Classification accuracy on group level was 85.6% for EOA, 63.5% for DCD and 91.2% for healthy control children. In comparison, using only the finger to nose test for classification 73.7% of EOA and 53.4% of DCD patients and 87.2% of healthy controls were accurately classified. For the ROC/precision recall plots the AUC was 0.96/0.89 for EOA, 0.92/0.81 for DCD and 0.97/0.94 for healthy control children.
Discussion: Using quantified movement information from all three SARA-kinetic upper limb tests improved the classification of all diagnostic groups, and in particular of the DCD group compared to using only the finger to nose test.

1. Introduction

Coordination is a complex ability which incorporates different parts of the body. Especially the cerebellum plays an important role in the execution of refined, coordinated movements and postural control [1]. Cerebellar dysfunction may induce ataxia which may include characteristics such as gait abnormalities and lack of balance control, insufficiently smoothly performed goal-directed movements, dysdiadochokinesia, dysmetria, overshoot, intention tremor, scanning speech and dysarthria, and cerebellar oculomotor disorders such as abnormal saccades and nystagmus [2,3]. Early onset ataxia (EOA) refers to a heterogeneous group of rare diseases (prevalence <1/1000 in children), causing ataxic features before the age of 25 years [4].

The clinical diagnosis of EOA is based on the visual assessment of clinically performed motor performances [5]. In addition to EOA, there are also other disorders, such as developmental coordination disorder (DCD) that may present with coordination impairment. Compared with EOA, DCD is a more commonly (prevalence 50–60/1000 children), non-progressive developmental disorder, that can be characterized by mild motor incoordination. This may cause difficulties in reaching motor milestones (such as grasping, sitting and standing) as well as problems with sensorimotor integration, postural control and visual-spatial planning [6,7]. By clinical definition, DCD is not attributable to any other underlying neurological condition, and should therefore be clinically discerned from rare EOA disorders. EOA and DCD patients share similar phenotypical

* Corresponding author. University medical Center Groningen (UMCG) Department of Neurology AB 51, PO Box 30.001, 9700, RB, Groningen, the Netherlands.
E-mail address: z.t.dominguez.vega@umcg.nl (Z.T. Dominguez-Vega).

https://doi.org/10.1016/j.ejpn.2021.07.009

characteristics, and because of this overlap differentiation between these two groups has been shown to be insufficiently accurate [5]. However, improvement of the diagnostic accuracy in the differentiation of EOA and DCD is important for choosing the correct diagnostic work-up, counselling surveillance, predicting familial recurrent risk, and treating the child.

The Scale for the Assessment and Rating of Ataxia (SARA) is one of the most frequently used rating scales to assess ataxia in the clinical environment [8]. The SARA protocol is evaluated separately for the upper and lower limbs, and the motor subscores (gait/posture; finger to nose, finger chase, alternating hand and heel shin slide) are determined by visual inspection of the child's motor performances and are also interpretable for age and/or immaturity of the child's motor system. For instance, during the SARA-finger to nose test, clinicians visually evaluate the decomposition of movement, dysmetria, kinetic and intention tremor. During the SARA-finger chase test, clinicians visually evaluate dysmetria by determining undershoot or overshoot of the movements towards the intended goal. During the SARA-fast alternating hand movements test, clinicians visually evaluate a-/dys-diadochokinesia and the speed and regularity of alternating hand movements [9]. In perspective of these semi-quantitatively visually assessed performances, it is understandable that essential features for diagnostic distinction between different diagnostic entities may be lost.

Use of inertial measurement units (IMU) has proven to be of help in the assessment of upper limb ataxic features and to distinguish patients with ataxia from other groups. In a study to identify and quantify the presence of ataxia in adults, Krishna et al. (2019) collected inertial data among others from the finger to nose and fast alternating movements tests. The authors aimed to extract features and to use linear discriminant analysis to classify healthy controls and groups of patients with either severity score one or two. By extracting information about acceleration, velocity and rotation, they found that acceleration was an important feature explaining performance on the fast alternating movements and heel-shin slide tests, whereas rotation was important for accurate classification of finger to nose test performance. Similarly, in two previous studies from our own group we extracted features from IMU signals obtained during finger to nose test performance to distinguish between EOA and DCD patients and healthy controls. We achieved 84% classification accuracy in distinguishing EOA patients and healthy controls, using only local curvature and instantaneous speed [11]. Comparing EOA and DCD patients and healthy controls, we correctly classified 87.4% of healthy controls, 74.4% of EOA and 24.8% of DCD patients by using different features such as intertrial variability, similarity of the movement trajectory to a straight line and principal component analysis to decompose the movements as input to train a random forest classifier [12].

Continuing on these previous studies which sought to quantify ataxia severity or to distinguish ataxia from other movement disorders, in this study we aim to combine information from all upper limb tests of the SARA, to improve the classification of EOA and DCD patients and healthy controls, compared to using only the finger to nose test. To do that, we first collected movement data using IMUs and then extracted features from the IMU signals for all three upper limb tests (finger to nose, finger chase and fast alternating hand movements). Features were selected based on their classification performance in earlier studies and their relevance for quantifying the most common problems a patient may present during clinical evaluation, such as dysmetria. For the finger to nose test we extracted features similar to those used by Martinez-Manzanera et al. (2018) and added local curvature and instantaneous speed in line with Aguilar et al. (2019). For the finger chasing test, we incorporated features that assess smoothness of movement [13], similar to the features used for the finger to nose test. For the fast alternating movements test we extracted movement speed as main feature. Finally, we trained a random forest classifier using leave one out cross validation to assess whether performance of the classifier in distinguishing between these three groups improved compared to our earlier work in which only one SARA test (finger to nose) was used for classification.

2. Methods

These data were collected as part of a larger project aiming to quantify the SARA performances in children with clinical coordination disorders, that were partially analysed in previous studies concerning the assessment of the SARA finger to nose test, alone [11,12]. The study was conducted in accordance with the principles of the Declaration of Helsinki (2013) and the research and integrity codes of the University Medical Center Groningen (UMCG). The Medical Ethical Committee of the UMCG provided a waiver for ethical approval since the SARA test battery was performed as part of clinical routine, and the attachment of IMUs to the skin is considered non-invasive. We obtained signed informed consent from the parents and informed assent from minors.

2.1. Participants

Patients were recruited from the outpatient clinic of the UMCG in the Netherlands between 2014 and 2019. Inclusion criteria were: children fulfilling the official criteria for: 1/EOA [14], 2/DCD [15], or 3/healthy controls (i.e., absence of a clinical neurologic or orthopaedic diagnosis that could theoretically interfere with coordinated motor performances). As the quantification of ataxic features is also dependent on the maturation of the central nervous system (i.e., the age of the child), we tried to match the age of the healthy controls with the age of the EOA and DCD patients. None of the included children received medication with known negative side-effects on motor coordination. Exclusion criteria were insufficiently performed SARA kinetic upper limb performances, as defined by less than 10 executed cycles of each SARA subscore task (see also assessments).

2.2. Clinical assessment

For the present study, we recorded all three SARA kinetic upper limb performances using IMUs, which concerns the SARA finger to nose, finger chase and fast alternating hand movement test performances. During the finger to nose test, participants are asked to touch the tip of their nose with the index finger and repeatedly move the index finger to a specific fixed point and back to the nose. In clinical routine assessments this fixed point is the index finger tip of the clinician. With the aim of making the test reproducible between participants, we asked children to perform a similar test by pointing to a fixed point (dot) on a computer screen, that was placed in front of the child at a distance of approximately 90% of maximum reach (see also Martinez-Manzanera et al. (2018)). This movement was repeated 10 times. The SARA finger chase test was similarly executed by pointing the finger to a moving target (dot) that changed position on the screen. The finger chasing test is similar to the finger to nose test, except that the target point changes position at every movement. We created a sequence of 15 moving points which change position every time the patient touches a point on the screen. This movement was repeated 15 times. The sequence was the same for all participants. Finally, during the SARA fast alternating movements test, children were asked to lay their hand on their lap while sitting comfortably, and then to perform 15 cycles of repetitive alternation of pro- and supinations of the hand on his/her thigh as fast and as precise as

possible. The number of actual executions depended on the child's physical ability to perform the test. Analogous to the SARA score instructions (i.e., the calculation of the average score of both sides), we pooled movements from both sides in our analysis. Patients with less than 10 movements (right and left arm combined) on any of the SARA kinetic subscore tasks were excluded for further analysis. Also, we videotaped the execution for clinical assessment of the EOA and DCD patients by one paediatric neurologist (author DAS) who provided subscores for each test and patient, in accordance with the official guidelines [16]. The average across right and left side was reported. We then analysed these data for potential differences between groups using (nonparametric) Kruskal-Wallis H tests.

2.3. Data acquisition

During performance of the finger to nose and finger chasing SARA tests, the participants were fitted with three light weight inertial measurement units (IMUs; Shimmer3, Shimmer, Dublin, Ireland); one on the upper arm, one on the forearm and one on the dorsal side of the index finger. Details of positioning can be found in [12]. For the fast alternating movements test, we changed the setup by removing the sensor on the index finger and changing the position of the shimmer on the forearm to the dorsal side of the wrist, just proximal of the wrist bones.

Before data acquisition in each participant, the three shimmer3 IMUs were programmed using Consensys v1.2.0 (Shimmer sensing, Dublin, Ireland) with LogAndStream_Shimmer3 v0.8.0 software for Bluetooth communication. Subsequently, each device was calibrated using Shimmer 9DoF Calibration v2.3 software. Calibration prevents misalignment of the electronic board containing the inertial sensors with the outer case and ensures proper recording of sensors. Once calibrated, the IMU was reprogrammed with SDLog_Shimmer3 v0.13 to log data via a Bluetooth protocol and finally the IMU was configured to use the triaxial ±4G (1G = 9.81 m/s^2) accelerometer, triaxial ±500 dps gyroscope and triaxial ± 1.9 Ga magnetometer and the sampling rate was set to 50 Hz. Data were collected via Bluetooth communication and stored in separate directories per patient, test and upper limb tested (right or left).

2.4. Signal preprocessing

For each patient we obtained six files (3 per arm), with inertial data. Each file contained information from the 3-axis accelerometer, 3-axis gyroscope, 3-axis magnetometer and quaternion data. The latter data were obtained with an implementation of the Madgwick filter [17] by Shimmer sensing in LabView (Austin, Texas, USA).

Position data were obtained from an upper limb model described in a previous study from our group [12]. Briefly, we built a 3D model of the upper limb in LabView which uses the quaternion data from the Madgwick filter. These data were then converted into angles which feed three rotational elements emulating the joints of the arm (shoulder, elbow and wrist). Rigid body elements were used to connect each joint to the next one; these elements represent the upper arm, fore-arm and hand, respectively. For the SARA finger to nose and finger chasing tests, we determined the positional data from the tip of the index finger of each participant, which is designated as END_POINT in our model. For analysis of the SARA fast alternating hand movements test, we determined the angular velocity using the IMU that was attached to the forearm.

With the aim to evaluate each movement separately we subsequently segmented the data from the three tests. A previous study from our group indicated that ataxic features from target to nose and vice versa are inhomogeneously performed [12]. To compensate for this, we separately evaluated each individual movement, for the finger to nose, the finger chasing as well as for the fast alternating movements (i.e., each pro- and supination) test. Accordingly, we segmented all movements for each test.

2.4.1. Finger to nose data segmentation

To segment finger to nose test data, we first identified the spatial axis with maximum movement variability using the interquartile range in the positional data. This axis may be different between patients. Then, a moving average filter with 15 samples per window was used to smooth the signal, and finally a peak detection algorithm was used to identify the start and end points of each separate movement as peaks or valleys in the signal. With the aim to *automatically* identify between nose to target and target to nose movements, which differs from the approach taken in Martinez-Manzanera et al. (2018), we used the positional data to calculate the Euclidian distance between the index finger and the shoulder (origin of the reference frame), employing that small distances indicate that the index finger is near the shoulder (i.e., that the index finger is on the tip of the nose) and larger distances indicate that the index finger is far from the shoulder (i.e., near the screen). Once each movement was classified as being either a nose to target or target to nose movement, we used 3D linear interpolation to ensure each movement trajectory was composed of exactly 100 points, allowing to directly compare between trajectories.

2.4.2. Finger chasing data segmentation

We identified individual movements in the finger chasing test, by using angular velocity data. First, we selected the signal from the shimmer attached to the upper arm and calculated the Euclidian norm for all data points. We selected this signal as it better represents the movement from one point to another because small movements such as isolated movements from the index finger are not visible. We then applied a moving average filter with 30 samples and a peak detection algorithm to identify the changes in direction. Identified extrema were visually verified. We used the location of peaks and valleys in the signal to segment the data, considering that each peak and its adjacent valleys represent a single movement. We therefore segmented the signal from valley to valley around an identified peak into separate trajectories per arm and participant.

2.4.3. Fast alternating movements data segmentation

The movements performed in this test are best described from an angle perspective since they concern rotations of the forearm. We thus used the angular velocity from the IMU attached to the wrist to identify and segment individual movements. To do this we first identified the data with the highest variance from the three spatial axes. We then applied a moving average filter with 15 samples per window and a peak detection algorithm to identify peaks and valleys in the signal. Identified extrema were visually verified. We also identified zero crossing points. We then segmented each part of the signal between two zero crossing points with a peak or valley in between as an individual movement. Finally, we identified positive segments of the signal as protonation movements and negative segments as supination movements, thereby taking into account the orientation of the sensor attached to the wrist.

2.5. Feature extraction

We based our feature extraction strategy on the known differences in smoothness and regularity of movement between EOA, DCD and healthy age-matched controls, while taking into account the results of our previous studies [11,12]. Details are provided in

the Appendix.

2.6. Classification

We used 34 features (18 for finger to nose, four for finger chase and 12 for fast alternating movements; for details, see Appendix) to quantify the lack of coordination in ataxic movements. We made a script in python version 3.7 in combination with scikit-learn version 0.32 to apply random forests as classification technique. We first created a table containing the extracted features, where each column contains values for a specific feature (34 columns) and each row contains combined information from the three tests. One row of this table contains information from two movements (n2t and t2n) of the finger to nose test, from one movement of the finger chasing test and two movements (protonation and supination) from the fast alternating movements test, randomly chosen; thereby building a 'combined movement' per patient, for classification. Each patient was represented by 10 rows in this table.

We decided to use the random forests (RF) classifier since it allows to gain insight in the most relevant features for classification, thereby contributing to 'explainable AI', i.e., providing input to clinicians regarding potentially novel characteristics of movement that are relevant for recognizing ataxia. Random forest is an ensemble technique which uses decision trees and averaging to improve the predictive performance [18]. The RF classifier creates a set of *n* decision trees and a subset of *m* features from the dataset; this splitting of decision trees and features has been shown to improve classification performance without overfitting.

A disadvantage of using random forests in combination with our dataset is the problem of unbalanced classes (EOA = 22, DCD = 14, CTRL = 24). One way to overcome this problem is the implementation of synthetic over and under sampling techniques which have been shown to have better classification performance [19]. Here, we use an adaptive synthetic sampling technique for oversampling the minority classes (EOA and DCD), thereby obtaining a balanced dataset [20].

Next, leave one patient out was used instead of the standard leave one instance out approach. We systematically took the 10 combined movements from the same patient out and used this information to test the performance for classification of new data. We collected the classification for each of these ten combined movements and then applied a majority vote strategy for general assessment of the performance of the classifier on the full dataset.

We used the features importance attribute of the RF classifier to quantify the relevance of each feature for classification.

Similar to a previous study from our group, we used 300 trees to build the random forest and used the Gini index as the separation criterion on each node to decrease the node impurity [12]. The RF algorithm uses an ensemble of decision trees where, on each iteration, a random subset of the features from the training set is taken. Therefore, this process can produce different results on every iteration. Classification was repeated 100 times and the average and standard deviation of the classification metrics were used to obtain a better estimate of classification performance, similar to Martinez-Manzanera et al. (2018).

Since we are dealing with an imbalanced dataset (EOA = 22, DCD = 14, CTRL = 24), we here implemented metrics to evaluate the performance of the classifier which take into account this imbalance: The Area Under the Receiver Operating Characteristic Curve (ROC-AUC) as well as precision and recall. A high ROC-AUC represents both high recall and high precision, where high precision indicates a low false positive rate, and high recall a low false negative rate. High scores for both show that the classifier is returning accurate results (high precision), as well as many true positive results (high recall).

Table 1
Participant characteristics for each of the three groups. EOA: early onset ataxia, DCD: developmental coordination disorder, CTRL: healthy control.

	EOA (n = 22)	DCD (n = 14)	CTRL (n = 24)
Age (median, iqr)	13 (6.5)	10 (1.5)	10.5 (4.2)
Range (years)	4–21	7–13	5–25
Sex (male/female)	13/9	9/5	9/15

Finally, we compared the classification accuracies obtained using the data from the three upper limb SARA tests against the results obtained using only the finger to nose SARA test, using McNemar's test [21]. To allow this comparison we also classified all participants using only the finger to nose test, using the preprocessing approach and features as detailed in Sections 2.4.1 and A.1.

3. Results

3.1. Participants

We acquired data from 79 participants. After applying all exclusion criteria, 60 participants remained; 22 EOA patients, 14 DCD patients and 24 healthy control participants. Age ranged from 4 to 21 years for the EOA group, from 7 to 13 years in the DCD group and from 5 to 25 years in the healthy control group. Age was not normally distributed for the three groups. For median and interquartile range, see Table 1. Age did not significantly differ between groups (Kruskal-Wallis test, $\chi2(2) = 2.8238$, $p = 0.2437$).

DCD patients had significantly lower SARA subscores than EOA patients for the finger to nose ($\chi2(2) = 11.5331$, $p = 0.0006$) and finger chasing tests ($\chi2(2) = 3.9889$, $p = 0.0457$) but groups did not differ on the fast alternating movements test ($\chi2(2) = 1.0507$, $p = 0.3053$).

3.2. Classification

The results of classification using leave one patient out over 100 iterations are illustrated in Fig. 1.

The estimated mean accuracy of the classifier on new data was 81.0% (5.1%). Table 2 depicts the confusion matrix obtained from 100 iterations. Classification accuracy on group level was 83.9% (SD 5.8%) for the EOA group, 62.1% (SD 5.8%) for the DCD group and

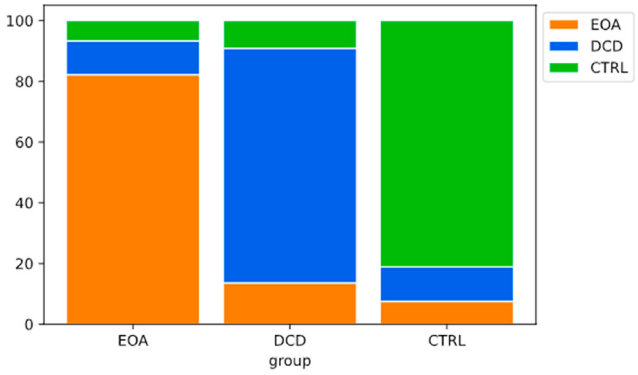

Fig. 1. Stacked bar plot presenting the percentages for classification over the 100 iterations where each bar represents 100% of the cases for each group, and how they are classified. For easy identification, we plotted DCD patients in blue, EOA patients in orange and healthy controls in green, in this and other figures. EOA: early onset ataxia, DCD: developmental coordination disorder, CTRL: healthy control. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Confusion matrix of the averaged 100 iterations. Overall mean accuracy is 81%. EOA: early onset ataxia, DCD: developmental coordination disorder, CTRL: healthy control.

	Prediction		
Actual	EOA	DCD	CTRL
EOA	83.9% (5.8%)	7.0% (3.4%)	9.0% (4.3%)
DCD	18.2% (4.2%)	62.1% (5.8%)	19.6% (5.4%)
CTRL	5.5% (3.5%)	3.7% (3.1%)	90.7% (5.1%)

90.7% (SD 5.1%) for the healthy control group. On average, 9.0% (SD 4.3%) of patients diagnosed as EOA were classified as healthy participants and 7.0% (SD 3.4%) were classified in the DCD group. Of the DCD patients on average 18.2% (SD 4.2%) were classified as EOA and 19.6% (SD 5.4%) as healthy controls. Finally, for healthy controls, on average 5.5% (SD 3.5%) were classified in the EOA group and 3.7% (SD 3.1%) were classified in the DCD group.

3.3. Feature importance

One of the main advantages of the RF classifier is that it allows to determine the most important features. In Fig. 2 the 20 most relevant features used by the classifier to discriminate between the three classes are displayed. This figure illustrates that including features from the finger to nose and the fast alternating movements tests is of most relevance for the classifier. The five features with largest importance were supination_pc1, t2n_pc1+pc2, t2n_s, n2t_s and supination_euM with a feature importance (averaged across 100 iterations) of 5.5%, 4.5%, 4.5%, 3.8%, 3.7% and 3.6%, respectively. Note that the only finger chase feature contributing in this list is at the last position.

3.4. Classification performance

Fig. 3 depicts the ROC-AUC plot to illustrate classification performance. The highest AUC value of 0.97 is found for the healthy participants, meaning that our classifier is best at classifying this group. EOA patients are also classified well with an AUC value of 0.96. The DCD patients are most difficult to classify with an AUC value of 0.92.

Fig. 4 presents the Precision-Recall plot. According to this representation of the classification results, the healthy participants are also classified best (AUC value of 0.94), followed by the EOA patients (AUC value of 0.89) and the DCD patients with an AUC value of 0.81.

Table 3
Median values across 100 contingency tables. Classifier1 is the classifier using the data from the three upper limb tests. Classifier2 is the classifier using data from the finger to nose test only.

	Classifier2 Correct	Classifier2 Incorrect
Classifier1 Correct	40	10
Classifier1 Incorrect	5	6

Next, we plotted the probabilities of being classified as EOA patient, DCD patient or healthy participant averaged across the 100 classification iterations, for every participant in the study. From Fig. 5 we can derive that control participants (green dots) are best separated (left upper corner) and just few participants score below the 0.6 probability of being a control participant. We can also see that the EOA patients (orange dots) are mostly gathered in the left lower corner and just one patient has more than 0.6 probability of being a DCD patient. Only 8 of the DCD patients are gathered in the lower right corner, with 2 patients having more than 0.6 probability of having EOA, 1 patient having more than 0.6 probability of being healthy and three DCD patients score in between the three groups (intersection region).

Finally, to perform McNemar's test we first determined the contingency table (see Table 3) for each iteration followed by calculating the mean values over 100 iterations, both for the classifier using the data from the three upper limb tests and the classifier using only the finger to nose test, Accuracies for the latter classifier were Classification accuracy on group level was 73.7% (SD

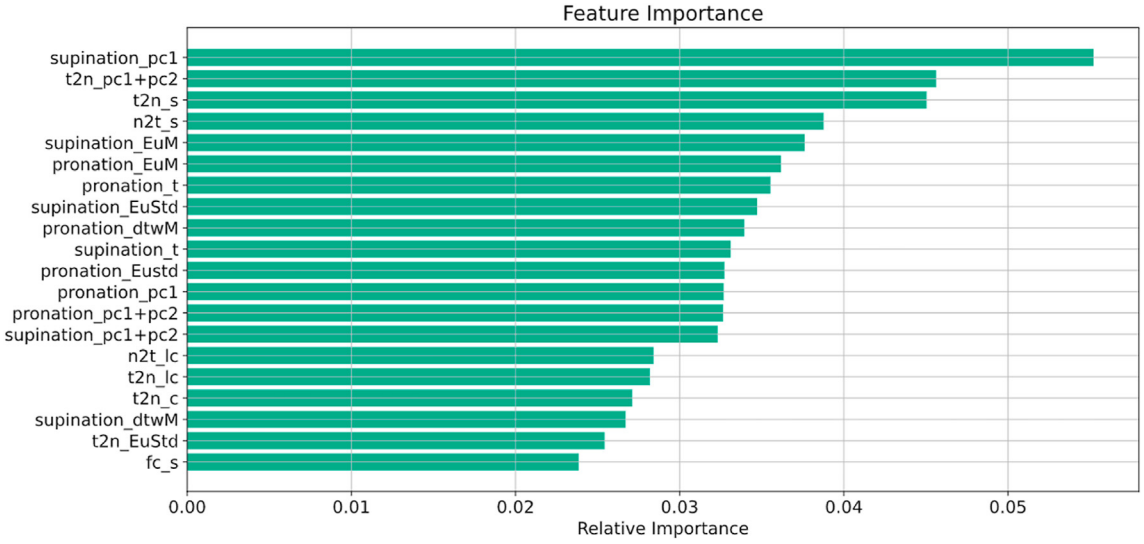

Fig. 2. Average relative feature importance of the 20 most relevant features for classification. The feature importance according to the decrease of node impurity is normalized across all features. For feature names please refer to the Appendix.

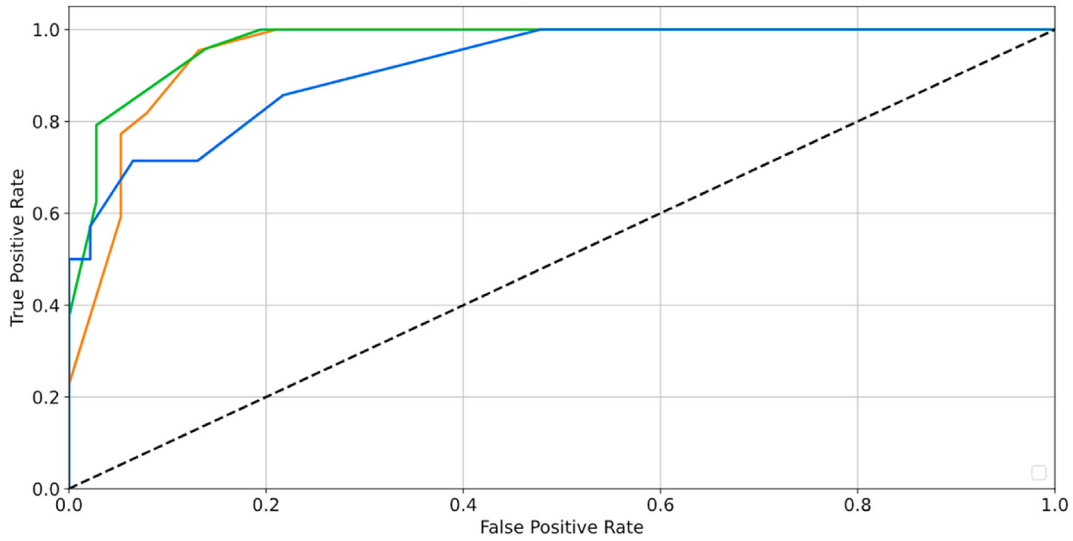

Fig. 3. Receiver Operating Characteristic (ROC) curves resulting from 100 iterations of random forests classification. Each line represents a group: EOA (orange), DCD (blue), and CTRL (green). EOA: early onset ataxia, DCD: developmental coordination disorder, CTRL: healthy controls. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

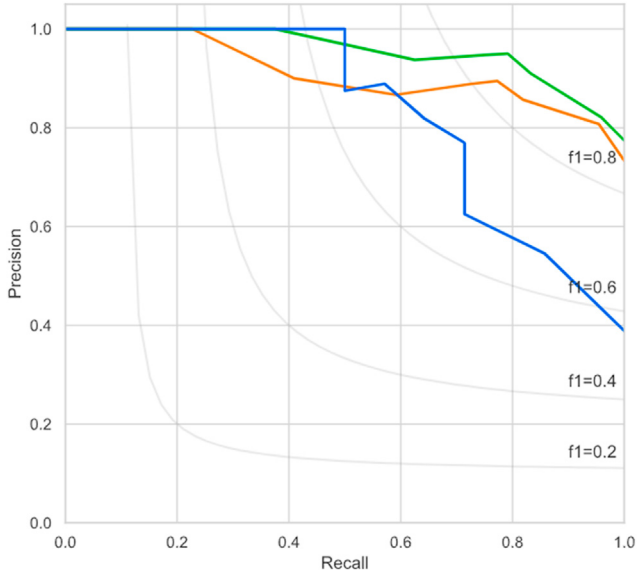

Fig. 4. Precision-Recall plot resulting from 100 iterations of random forests classification. Each line represents a group: EOA (orange), DCD (blue), and CTRL (green). EOA: early onset ataxia, DCD: developmental coordination disorder, CTRL: healthy participants. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

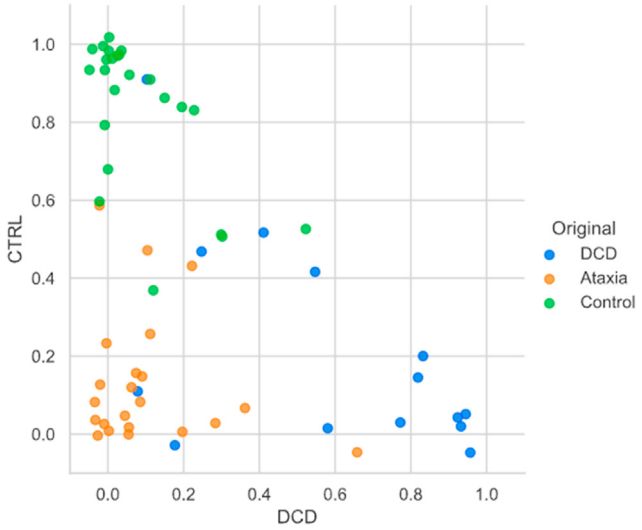

Fig. 5. Average probability of being classified as EOA patient, DCD patient or healthy control across the 100 iterations of classification. EOA: early onset ataxia, DCD: developmental coordination disorder, CTRL: healthy control. A dot in the upper left corner of the plot indicates that the participant is always classified as healthy control, a dot in the lower right corner indicates that the participants is always classified as DCD patient and a dot in the lower left corner indicates that the participant is always classified as EOA patient. The central area of the plot is the intersection region where a participant is variably classified as belonging to one of the three groups, e.g. a dot at position (0.3,0.2) indicates that this participant has 0.3 probability of being classified as a DCD patient, 0.2 probability of being classified as a healthy participant and (1−0.3-0.2=)0.5 probability of being classified as an EOA patient. For better visualization of the position of each participant in this plot, we added jitter to the exact values; this explains the slightly negative probability values.

8.2%) for the EOA group, 53.4% (SD 10.6%) for the DCD group and 87.2% (SD 5.4%) for the healthy control group.

Thus, there are six cases (median) incorrectly classified by both classifiers. Using all three upper limb tests, five (median) more cases are classified correctly, compared to using only the finger to nose test. Results from McNemar's test (statistic = 5.0, p = 0.302) indicate that the improvement in classification using three upper limb tests instead of one is not significant, however.

4. Discussion

In this study we combined information extracted from IMU data obtained while children were performing three SARA kinetic upper limb tests (finger to nose, finger chasing, fast alternating movements), and showed that this allows for improved classification of EOA and DCD patients and healthy controls compared to earlier studies employing only one (finger to nose) test. The improvement in classification was not significant, however, which may be due to the relatively small sample. We used features related to smoothness and velocity of movement since they are considered relevant for the correct identification of ataxia in the clinical environment.

We included 60 patients who were all able to perform the SARA tests, in particular for assessment of the upper limb, with a minimum of 10 movements when both sides were combined (right and left). Using the extracted features, we trained a random forest classifier, combined with a synthetic oversampling technique, to differentiate between EOA and DCD patients and healthy controls. As the combination of the three upper limb tests improved the classification of these three groups, compared to earlier studies, we hope that this information may be useful for clinicians as an aid in the identification of ataxic movements in children. The features that we identified as most relevant for classification may also be useful in future studies investigating the application of IMUs to monitor patients in their home environments for the effects of disease progression or interventions.

In the present study, in which we used information from three upper limb SARA tests, we found that on average 90.7% of healthy controls, 83.9% of EOA and 62.1% of DCD patients were correctly classified. In a previous study from our group [12], in which only the finger to nose test was used, we found that on average 87.4% of healthy controls, 74.4% of EOA and 24.8% of DCD patients were correctly classified. Repeating the classification using only the information from the finger to nose test but with our new methodology and data we found that on average 87.2% of healthy controls, 73.7% of EOA and 53.4% of DCD patients were correctly classified. The improved classification accuracy using more tests was anticipated, both from a theoretical point of view – using more information may result in better classification - but also from a clinical point of view, where diagnosis is also based on multiple tests. In addition, we now also used a synthetic oversampling technique to deal with the imbalanced dataset instead of giving different weights to the classes, as was done in our previous study, which is one of the reasons why the classifier using the finger to nose test data only in the current study performed better that the one in our previous study [12]. Finally, in the present study we included more patients compared to our previous study; we increased the sample with 13 EOA and seven DCD patients, as well as eight healthy controls. Comparing the present IMU results derived from SARA kinetic upper limb data alone, with the percentage of unanimous clinical assessments made by three movement disorder specialists based on the complete SARA (SARA -gait/posture, -kinetics and -speech), also revealed better results using the digital kinetic IMU-derived data (83,9% and 62.1% versus 73% and 20%; for EOA and DCD patients, respectively [5]. This may implicate that the presently applied technique could be a worthwhile instrument to be used for the classification of coordination impairment.

For the current classification, the most relevant feature was the first principal component of the supination movement (supination_PC1), whereas in our previous study this was the first principal component of the target to nose trajectories (t2n_PC1). More generally, for the current classification, the most relevant features were all related to the fast alternating movements test and the finger to nose test. In particular supination_pc1, t2n_pc1+pc2, t2n_s, n2t_s and supination_EuM were the top five most relevant features, A feature derived from the finger chasing test only enters the list of most relevant features at position 20. This could be due to the fact that for the finger chasing test we only extracted four features compared to 18 for the finger to nose and 12 for the fast alternating movements test. However, we also noticed that task execution and recording for the finger chasing test were less reliable for several patients, because some patients touched the screen not with just their finger tip, but also with parts of other fingers.

Here, we used 34 features to obtain the high classification accuracy. This may not always be necessary, depending on the exact classification task: in another study from our group [11], where we tried to distinguish between healthy controls and movement disorder patients (Adult Onset Ataxia, EOA and DCD combined) we obtained 84% classification accuracy, using only two features (local curvature and instantaneous speed, as derived from the SARA finger to nose test, for both t2n and n2t movements). We here obtained a similar classification accuracy for distinguishing EOA from the other two groups (83.9%) and a higher accuracy when distinguishing healthy controls from the other two groups (90.7%). Of course, we here also distinguished DCD patients from EOA patients and healthy controls with reasonable accuracy (62.1%), which is a much more challenging task, with DCD patients phenotypically overlapping with both EOA patients and healthy controls. On the other hand, the features used by Aguilar et al. (2019) were also included for our current classification (as n2t_lc, t2n_lc, n2t_s, t2n_s) and yet, we need 28 more features to achieve our goal. However, the local speed features are in the top 5 of most relevant features and the local curvature features also remain in the top 20.

To the best of our knowledge there is only scarce literature on the use of inertial sensors for classification of ataxia patients and healthy controls. In a similar study, Krishna et al. (2019) tried to evaluate disability due to cerebellar ataxia in adult cerebellar ataxia patients and healthy controls by predicting SARA severity scores (divided into control scores and low and high severity patient scores), for three SARA tests (finger to nose, fast alternating movements and heel to shin). In a similar approach as we used, the authors from this study tried to classify the three severity groups, obtaining average accuracy values of 0.92 and 0.93 for the finger to nose and fast alternating movements test, respectively. To compare their results with ours, we calculated the accuracy of classifying EOA and healthy controls using the information in the confusion matrix, resulting in accuracy values of 0.89 (EOA) and 0.95 (healthy controls), thereby using information from both tests. Our results thus seem in general agreement to those presented by Krishna et al. (2019) although our goal was different.

Also, according to Krishna et al. (2019) acceleration is a major feature for predicting fast alternating movements and heel to shin test scores, whereas rotation is the main feature responsible for predicting finger to nose test scores. In our case, we found that angular velocity information from the fast alternating test in combination with positional information from the finger to nose test gives the best prediction according to the node impurity in the random forest classifier.

To the best of our knowledge, this is the first study combining the information from the three upper limb SARA tests, to extract meaningful features and classify EOA and DCD patients and healthy children. Furthermore, our approach to automatically identify and segment individual movements increases reproducibility of this study. In the studies mentioned here [10–12], identification of movements and segmentation was done (semi-)manually which reduces the reliability of feature extraction.

This study has some limitations. We are aware that a larger number of patients could better represent the population. However, as explained, EOA is a rare diagnosis and the presently studied patient cohorts have been carefully collected and matched over the last years at a tertiary university clinic, specialized in movement disorders. As a tertiary center, we are aware that we could only include a limited number of DCD patients. However, in the future we aim to conduct a second, collaborative study with a more extensive inclusion of patients fulfilling the DCD criteria. Also, the number of sensors as well as their attachment could be improved if smaller and lighter sensors would be available. This could improve the performance of patients during the tests, reducing the number of patients that have to be left out of data analysis (currently 40 out of a 100). Future work could also include the assessment of longitudinal data points and the association of our IMU outcomes with functional data. Finally, we are aware that combining IMU data

from all SARA motor tests (including SARA -kinetic subscores from the lower limb and SARA −gait/posture) could even further improve the accuracy of the resulting diagnostic classification.

5. Conclusion

Combined IMU data from the three upper limb SARA tests (finger to nose, finger chasing and fast alternating movements) provided better EOA and DCD classification, compared to using data from the SARA finger to nose test alone and also compared to clinical phenotype classifications (based on the total SARA). Furthermore, automatically identified quantified movement features, have the advantage of better reproducibility than clinically phenotyping by gestalt perception alone. In particular, the fast alternating movements test added relevant information for the classification of these three groups, whereas the finger chasing test did not. Although, automatic classification cannot replace the complexity of a full clinical neurologic assessment and investigation at the outpatient clinic, the current results may implicate that this technique could provide a worthwhile instrument for clinical diagnostic support.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

none.

Appendix

A.1 Finger to nose

We extracted features from the finger to nose test data employing the features from Martinez-Manzanera et al. (2018) and Aguilar et al. (2019). We here briefly repeat how we extracted these features. Using the positional data, we first calculated the variances in these data explained by the first and by the first two principal components (PC1 and PC1+PC2). The span of the first two principal components is a plane that best fits the three-dimensional movement trajectory. As we expect that the movement trajectories of patients are less smooth than those of healthy participants, the variance explained by the first and the first two principal components should be larger for the healthy participants. We defined these features as n2t_pc1, n2t_pc1+pc2, t2n_pc1 and t2n_pc1+pc2, where n2t refers to nose to target and t2n to target to nose movements.

Because their curves are less smooth and more irregular, the curvature of the trajectories in the finger to nose test is supposedly higher in patients than in healthy participants [11,12], which is why we also employed three features related to curvature. First, we determined the projection of the 3D movement trajectory on the first two principal components and obtained the dynamic time warping (DTW) distance to a Bezier curve generated with three points (start, end and middle point of the original trajectory). For the second curvature feature, we determined the DTW distance of the original movement trajectory to a straight line generated from the starting to the end point of the 3D trajectory. For the third curvature feature, we used local curvature defined as the inverse of the radius of a circle fitting through three consecutive points of the original 3D movement trajectory and averaged it across the trajectory. These features were defined as n2t_c, t2n_c, n2t_l, t2n_l, n2t_lc and t2n_lc, respectively.

Variability across movement trajectories was also expected to be larger in the patient groups than in the healthy participants. To evaluate inter-movement trajectory variability, we therefore defined a further three features. First, we calculated the mean trajectory across movements per participant. Then we calculated the Euclidean distance from the mean trajectory to each point on the individual trajectories and the mean and standard deviation of these distances. Finally, we calculated the mean of the DTW distance from individual trajectories to the mean trajectory to investigate the similarity among trajectories of the same test execution. We identified these features as n2t_EuM, n2t_EuStd, t2n_EuM, t2n_EuStd, n2t_dtwM and t2n_dtwM.

Finally, we calculated the instantaneous speed which in combination with local curvature has shown good classification performance when distinguishing between EOA and healthy participants [11]. Instantaneous speed was calculated using two consecutive points of the trajectory, according to the method of Aguilar et al. (2019) and also averaged across the trajectory. We identified instantaneous speed as n2t_s and t2n_s. In total, we thus extracted nine features for each movement (n2t or t2n) from the finger to nose test.

A.2 Finger Chasing

Similar to the finger to nose test, we expected the patients to have more irregular movement trajectories than healthy participants. To evaluate regularity of movement, we thus again calculated the first two principal components explaining most of the variance in each finger chasing movement trajectory. We therefore defined the features fc_pc1 and fc_pc1+pc2 similar to those for the finger to nose test, to represent regularity of individual finger chasing movement trajectories.

In this test the participant is expected to go from one point to the next as fast as possible. We here expect healthy participants to have shorter and faster movements compared to EOA or DCD patients. We therefore again calculated instantaneous speed (fc_s), averaged across the trajectory, to evaluate the velocity during each finger chasing movement trajectory. Finally, we calculated local curvature (fc_lc), averaged across the trajectory, expecting that healthy participants will have smoother movements with smaller curvature values than EOA and DCD patients for which curvature is expected to be higher. Since the individual trajectories differ from each other by design of the test, we only calculated local curvature to evaluate smoothness of movement and not the other curvature features that were used for the finger to nose test. In summary, we calculated four features (for each movement) for the finger chasing test.

A.3 Fast alternating movements

Similar to the finger to nose and finger chasing tests, the trajectories during the fast alternating movements test are expected to be smoother and more regular in healthy participants compared to EOA and DCD patients. We here propose that principal component analysis may also be helpful to identify irregularity in angular velocity; similar to the finger to nose test we expect the variance explained by the first and the first two principal components to be higher in healthy participants than in EOA or DCD patients. We

therefore calculated explained variance for the first and the first and second principal components per movement (protonation or supination) to describe regularity of movement. We identified these features as protonation_pc1, protonation_pc1+pc2, supination_pc1 and supination_pc1+pc2, respectively. On the other hand, with the aim of evaluating the regularity of movement across trajectories for the same patient we calculated the mean and standard deviation of the Euclidean distance and the mean of the dynamic time warping (DTW) distance similar to what we did for the finger to nose test. In this case we used the angular velocity for protonation and supination movements, separately. We first used 3D linear interpolation again so that the segmented signal had 100 samples, then calculated the average trajectory, and subsequently calculated the individual distance (Euler and DTW) from each trajectory to the average angular velocity trajectory. Finally, we calculated the mean and standard deviation for the Euclidean distance and the mean for the DTW distance. We identified these features to be protonation_EuM, protonation_EuStd, protonation_dtwM, supination_EuM, supination_EuStd and supination_dtwM.

We calculated the time spent during individual protonation or supination movements and identified this feature as protonation_t and supination_t. We thus extracted six features for each movement (protonation or supination) for the fast alternating movements test.

Contributors

ZTDV: Study design, acquisition and analysis of data, manuscript drafting. DD: Acquisition and analysis of data. JWJE: reviewing the manuscript, reviewing statistical methods. DAS: study design, reviewing of the manuscript. NMM: Study design, article drafting, reviewing of the manuscript, reviewing statistical methods.

Funding

The Mexican National Council for Science and Technology (CONACYT) provided ZTDV with a scholarship to financially support this research. Funding sources were not involved in study design, collection, analysis and interpretation of data, writing of the report, or in the decision to submit the article for publication.

Disclosure

All authors have approved the final version of this article.

Declarations of interest

None.

Patient consent for publication

Obtained.

Features and explanation.

	Feature	Description
1	n2t_pc1	Variance explained by the first principal component in the position data from the index finger and during the nose to target trajectory.
2	n2t_pc1+pc2	Variance explained by the first and second principal components in the position data from the index finger and during the nose to target trajectory.
3	t2n_pc1	Variance explained by the first principal component in the position data from the index finger and during the target to nose trajectory.
4	t2n_pc1+pc2	Variance explained by the first and second principal components in the position data from the index finger and during the target to nose trajectory.
5	n2t_c	Dynamic time warping distance to a Bezier curve generated with three points (start, end and middle point of the original trajectory) and during the nose to target trajectory.
6	t2n_c	Dynamic time warping distance to a Bezier curve generated with three points (start, end and middle point of the original trajectory) and during the target to nose trajectory.
7	n2t_l	Dynamic time warping distance of the original movement trajectory to a straight line generated from the starting to the end point of the 3D trajectory during the nose to target trajectory.
8	t2n_l	Dynamic time warping distance of the original movement trajectory to a straight line generated from the starting to the end point of the 3D trajectory during the target to nose trajectory.
9	n2t_lc	Mean local curvature calculated from the positional data during the nose to target trajectory.
10	t2n_lc	Mean local curvature calculated from the positional data during the target to nose trajectory.
11	n2t_EuM	Mean Euclidean distance from the mean trajectory to each point on nose to target trajectories.
12	n2t_EuStd	Standard deviation of Euclidian distance from the mean trajectory to each point on nose to target trajectories.
13	t2n_EuM	Mean Euclidean distance from the mean trajectory to each point on target to nose trajectories.
14	t2n_EuStd	Standard deviation of Euclidian distance from the mean trajectory to each point on target to nose trajectories.
15	n2t_dtwM	Mean dynamic time warping (DTW) distance from the mean trajectory to each point on the nose to target trajectory.
16	t2n_dtwM	Mean dynamic time warping (DTW) distance from the mean trajectory to each point on the target to nose trajectory.
17	n2t_s	Mean instantaneous speed during the nose to target trajectory.
18	t2n_s	Mean instantaneous speed during the target to nose trajectory.
19	fc_pc1	Variance explained by the first principal component in the position data from the index finger and during each finger chasing trajectory.
20	fc_pc1+pc2	Variance explained by the first and second principal components in the position data from the index finger and during each finger chasing trajectory.
21	fc_s	Mean instantaneous speed during each finger chasing trajectory.
22	fc_lc	Mean local curvature calculated from the positional data during each finger chasing trajectory
23	pronation_pc1	Variance explained by the first principal component in the angular velocity data of the wrist during the pronation trajectory.
24	pronation_pc1+pc2	Variance explained by the first and second principal components in the angular velocity data of the wrist during the pronation trajectory.
25	supination_pc1	Variance explained by the first principal component in the angular velocity data of the wrist during the supination trajectory.
26	supination_pc1+pc2	Variance explained by the first and second principal components in the angular velocity data of the wrist during the supination trajectory.
27	pronation_EuM	Mean Euclidean distance from the mean trajectory to each point on pronation trajectories.
28	pronation_EuStd	Standard deviation of Euclidian distance from the mean trajectory to each point on pronation trajectories.
29	pronation_dtwM	Mean dynamic time warping (DTW) distance from the mean trajectory to each point on pronation trajectories.
30	supination_EuM	Mean Euclidean distance from the mean trajectory to each point on supination trajectories.
31	supination_EuStd	Standard deviation of Euclidian distance from the mean trajectory to each point on supination trajectories.
32	supination_dtwM	Mean dynamic time warping (DTW) distance from the mean trajectory to each point on supination trajectories.
33	pronation_t	Time spent during individual pronation trajectories.
34	supination_t	Time spent during individual supination trajectories.

Ethics approval

Medical-ethical committee of the University Medical Center Groningen.

References

[1] D.E. Haines, E. Dietrichs, The cerebellum - structure and connections, in: Handbook of Clinical Neurology, 103, Elsevier B.V., 2012, pp. 3–36.

[2] E. D'Angelo, C.I. De Zeeuw, Timing and plasticity in the cerebellum: focus on the granular layer, Trends Neurosci. (2009) 30–40, https://doi.org/10.1016/j.tins.2008.09.007.

[3] M. Manto, J.M. Bower, A.B. Conforto, J.M. Delgado-García, S.N.F. Da Guarda, M. Gerwig, et al., Consensus paper: roles of the cerebellum in motor control-the diversity of ideas on cerebellar involvement in movement, Cerebellum 11 (2012) 457–487. Springer-Verlag.

[4] A.E.E. Harding, Classification of the Hereditary Ataxias and Paraplegias, 321, Elsevier, 1983.

[5] T.F. Lawerman, R. Brandsma, N.M. Maurits, O. Martinez-Manzanera, C.C. Verschuuren-Bemelmans, R.J. Lunsing, et al., Paediatric motor phenotypes in early-onset ataxia, developmental coordination disorder, and central hypotonia, Dev. Med. Child Neurol. (2019), https://doi.org/10.1111/dmcn.14355.

[6] J.G. Zwicker, C. Missiuna, L.A. Boyd, Neural correlates of developmental coordination disorder: a review of hypotheses, J. Child Neurol. 24 (10) (2009) 1273–1281, https://doi.org/10.1177/0883073809333537.

[7] P.H. Wilson, S. Ruddock, B. Smits-Engelman, H. Polatajko, R. Blank, Understanding performance deficits in developmental coordination disorder: a meta-analysis of recent research, Dev. Med. Child Neurol. 55 (3) (2013) 217–228, https://doi.org/10.1111/j.1469-8749.2012.04436.x.

[8] T. Schmitz-Hübsch, Scale for the assessment and rating of ataxia (SARA), Encycl Mov Disord (2010) 95–99, https://doi.org/10.1016/B978-0-12-374105-9.00534-7.

[9] F. Bodranghien, A. Bastian, C. Casali, M. Hallett, E.D. Louis, M. Manto, et al., Consensus paper: revisiting the symptoms and signs of cerebellar syndrome, Cerebellum (2016) 369–391, https://doi.org/10.1007/s12311-015-0687-3.

[10] R. Krishna, P.N. Pathirana, M. Horne, L. Power, D.J. Szmulewicz, Quantitative assessment of cerebellar ataxia, through automated limb functional tests, J. NeuroEng. Rehabil. 16 (1) (2019) 31, https://doi.org/10.1186/s12984-019-0490-3.

[11] V.S. Aguilar, O.M. Manzanera, D.A. Sival, N.M. Maurits, J.B.T.M. Roerdink, Distinguishing patients with a coordination disorder from healthy controls using local features of movement trajectories during the finger-to-nose test, IEEE Trans. Biomed. Eng. 66 (6) (2019) 1714–1722, https://doi.org/10.1109/TBME.2018.2878626.

[12] O. Martinez-Manzanera, T.F. Lawerman, H.J. Blok, R.J. Lunsing, R. Brandsma, D.A. Sival, et al., Instrumented finger-to-nose test classification in children with ataxia or developmental coordination disorder and controls, Clin Biomech 2018 60 (June 2017) 51–59, https://doi.org/10.1016/j.clinbiomech.2018.10.007.

[13] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, E. Burdet, On the analysis of movement smoothness, J. NeuroEng. Rehabil. 12 (1) (2015) 37–39, https://doi.org/10.1186/s12984-015-0090-9.

[14] A.E. Harding, Clinical features and classification of inherited ataxias, Adv. Neurol. 1–14 (1993).

[15] A. Wenzel, Diagnostic and Statistical Manual of Mental Disorders, fifth ed., (DSM-5). The SAGE Encyclopedia of Abnormal and Clinical Psychology, 2017.

[16] T. Schmitz-Hübsch, S.T. Du Montcel, L. Baliko, J. Berciano, S. Boesch, C. Depondt, et al., Scale for the assessment and rating of ataxia: development of a new clinical scale, Neurology 66 (11) (2006) 1717–1720, https://doi.org/10.1212/01.wnl.0000219042.60538.92.

[17] Madgwick, O.H. Sebastian, A.J.L. Harrison, R. Vaidyanathan, Estimation of IMU and MARG Orientation Using a Gradient Descent Algorithm 2011, IEEE, 2011, pp. 1–7.

[18] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[19] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, G. Bing, Learning from class-imbalanced data: review of methods and applications, Expert Syst. Appl. 73 (2017) 220–239, https://doi.org/10.1016/j.eswa.2016.12.035.

[20] H. He, Y. Bai, E.A. Garcia, S. Li, ADASYN: adaptive synthetic sampling approach for imbalanced learning, Proc Int Jt Conf Neural Networks (3) (2008) 1322–1328, https://doi.org/10.1109/IJCNN.2008.4633969.

[21] T.G. Dietterich, Approximate statistical tests for comparing supervised classification learning algorithms, Neural Comput. 10 (7) (1998) 1895–1923, https://doi.org/10.1162/089976698300017197.