



# Kinematic features of a simple and short movement task to predict autism diagnosis

**DOI:**  
[10.1109/EMBC.2019.8857307](https://doi.org/10.1109/EMBC.2019.8857307)

**Document Version**  
Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

**Citation for published version (APA):**  
Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. (2019). Kinematic features of a simple and short movement task to predict autism diagnosis. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* IEEE. <https://doi.org/10.1109/EMBC.2019.8857307>

**Published in:**  
2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)

**Citing this paper**  
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

**General rights**  
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Takedown policy**  
If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [uml.scholarlycommunications@manchester.ac.uk](mailto:uml.scholarlycommunications@manchester.ac.uk) providing relevant details, so we can investigate your claim.



# Kinematic features of a simple and short movement task to predict autism diagnosis

Andrius Vabalas, *Student Member, IEEE*, Emma Gowen, Ellen Poliakoff  
and Alexander J. Casson, *Senior Member, IEEE*

**Abstract**—Autism is a developmental condition primarily identified by social and communication deficits. However, over 70% of autistic individuals also show motor function deficits, which are evident even when simple stereotyped movements are performed. In this study, we have asked 24 autistic and 22 non-autistic adults to perform pointing movements between two markers 30 cm apart as quickly and as accurately as they can for 10 seconds. Motion tracking was employed to collect data and calculate kinematic features of the movement and aiming accuracy. At the group level, the results showed that autistic individuals performed pointing movements slower but more accurately compared to non-autistic individuals. At the individual level, we have used Machine Learning methods to predict autism diagnosis. Nested result Cross-Validation was used, which in contrast to commonly used K-fold Cross-Validation avoids pooling training and testing data and provides robust performance estimates. Our developed models achieved a statistically significant classification accuracy of 71% and showed that even a simple and short motor task enables discrimination between autistic and non-autistic individuals.

## I. INTRODUCTION

Autism Spectrum Condition (ASC) is a group of complex developmental conditions primarily characterised by social and communication deficits. Stereotyped and repetitive movements are also recognised as a symptom but receive little attention in identifying ASC [1]. Nonetheless, growing research interest for motor impairment in ASC demonstrated that balance, gait, praxis and visuomotor functions are implicated and deficits show large, highly significant effect sizes [2]. Even very simple movement tasks like reach to grasp [3], repetitive hand pronation-supination and finger tapping [4] show differences between ASC and typically developing (TD) individuals.

In this study we have asked adult ASC and TD participants to perform a simple and short movement task — to point between two points as quickly and as accurately as they can. The movements were objectively measured using motion tracking and Machine Learning (ML) methods were applied to predict ASC diagnosis.

The strength of ML is the ability to find complex interactions between multiple variables and this makes it well

A. Vabalas and A. J. Casson are with the School of Electrical and Electronic Engineering, The University of Manchester, UK. Email: {andrius.vabalas, alex.casson}@manchester.ac.uk

E. Gowen and E Poliakoff are with the School of Biological Sciences, The University of Manchester, UK. Email: {emma.gowen, ellen.poliakoff}@manchester.ac.uk

A. Vabalas was supported by the UK Engineering and Physical Sciences Research Council and its Doctoral Training Partnership with the University of Manchester.

suitable for identification of ASC biomarkers as ASC is a complex and heterogeneous condition which can have different expressions between affected individuals. Additionally, the current diagnostic process is long and subjective, based on observation, interview, and questionnaire techniques applied by clinical experts. The prospect of developing automated algorithms assisting with ASC identification would speed up the diagnostic process and make it more objective. Several previous studies have applied ML methods for ASC identification and a few also used kinematic data: tracking game-play with a sensors on a tablet screen surface [5], tracking reach-and-throw a ball in basket movements [6] and tracking a simple movement imitation task [7]. Those studies achieved high classification accuracy rates of 86.7% to 96.7%. However, these studies used K-fold Cross-Validation (CV), which was demonstrated to produce overoptimistic performance estimates [8], especially when a sample size is small [9]. It was also not assessed whether classification performance was statistically significantly different from random guessing.

In this study, we have explored whether kinematic characteristics of a short and easy to perform movement task can be used to help predict ASC diagnosis. We have developed automated feature selection and classification algorithms. Importantly, we have used nested result CV, which avoids pooling training and testing data, and gives “almost unbiased estimate” [8]. Additionally, we have assessed if classification results were statistically significantly different from random guessing. The remainder of this paper is organised as follows. In Section II we present the methods used for data collection and ML analyses. Section III presents behavioural and ML results. Finally, conclusions are drawn in Section IV.

## II. METHODS

### A. Experiment and data

24 ASC (9 female, age 31.5 years) and 22 TD (7 female, age 30.8 years) IQ matched participants performed a simple pointing task. Two 8 mm diameter red stickers were attached 30 cm apart on the horizontal surface of a table in front of seated participants who were instructed: “With your index finger point between these two targets as quickly and as accurately you can”. The task was performed with dominant and non-dominant hand. Polhemus Fastrak motion tracker was used for kinematic data collection with a single motion sensor attached to the distal phalange of the index finger. Movement was sampled at 120 Hz in X, Y, Z coordinates and features based on velocity, acceleration, jerk and amplitude were calculated for each pointing movement (Table I). In

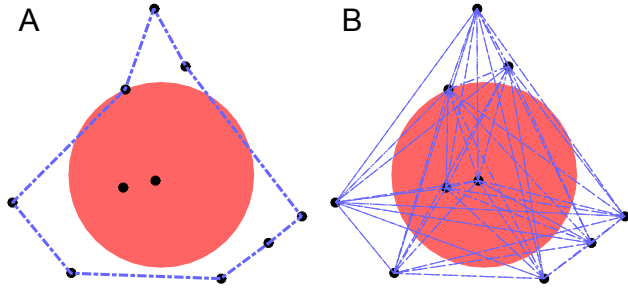


Fig. 1. Pointing accuracy measures, black points represent locations where participant touched a table surface when performing pointing to a target movement. (a) Area bounded by pointing locations. (b) Distances between individual pointing locations. Red area illustrates a target sticker.

TABLE I  
MOVEMENT MEASURES

1. Duration (s)	11. Peak deceleration (mm/s)
2. Peak velocity (mm/s)	12. Peak deceleration time (s)
3. Time to peak velocity (s)	13. Peak deceleration time (%)
4. Horizontal amplitude before peak acceleration (%)	14. Horizontal ampl. at which peak deceleration occurs (%)
5. Time before peak velocity (%)	15. Dimensionless jerk
6. Time after peak velocity (%)	16. Horizontal amplitude (mm)
7. Peak acceleration (mm/s)	17. Vertical amplitude (mm)
8. Time to peak acceleration (s)	18. Peak vertical ampl. time (s)
9. Time to peak acceleration (%)	19. Peak vertical ampl. time (%)
10. Horizontal amplitude at which peak acceleration occurs (%)	20. Horizontal ampl. of peak vertical ampl. (%)

addition, accuracy features were calculated based on the area bounded by points where the finger touched the horizontal surface when performing targeted movements to sticker locations (Fig. 1a) and by the average distance between all points (Fig. 1b). In total there were 60 features which were means and standard deviations (SDs) calculated by pooling movements performed with dominant and non-dominant hands. The experimental procedures involving human subjects described in this paper were approved by the University of Manchester research ethics committee, ref: 2017-2541-4204.

### B. Data preparation

Individual outliers for each participant were removed and group outliers (1.4% of all data points) were replaced with group means. Outliers were identified based on non-recursive procedure recommended by Van Selst and Jolicœur (1994) [10]. Features were normalised by transforming to z-scores.

### C. Classification algorithm

For classification Support Vector Machine (SVM) algorithm [11] was used. It separates the classes by maximising the gap between training examples from each class. To deal with non-linearly separable classes SVM uses kernel functions and penalty parameter  $C$ , which weighs the importance of misclassification. In this study for classification SVM with

Radial Basis Function (RBF) kernel was used. RBF kernel has a regularisation parameter  $\gamma$ , which regulates the spread of the kernel function and in turn determines the flexibility of the separating hyperplane. Regularisation parameters  $C$  and  $\gamma$  were optimised using grid search approach, which evaluates classification accuracy with different combinations of  $C$  and  $\gamma$  parameters by using CV. SVM and grid search were implemented with Scikit-learn [12] library.

### D. Feature selection

Before classification the ML algorithm pipeline included feature selection. Four feature selection methods were used to reduce feature space.

*SVM Recursive Feature Elimination (RFE)* [13] algorithm selected features based on their importance for SVM classifier to separate classes. In this study SVM-RFE started with a full feature set and in each iteration one feature was eliminated, which by an SVM algorithm was deemed least important to separate classes, using weight vector of dimension length(s) as a ranking criterion [13]. The Final feature set was selected from the iteration in which SVM achieved best classification performance.

*Students t-test (two-sample)* was used as a filter feature selection method. 10 features with the highest absolute value of the  $t$ -statistic and thus with most different means between two classes were selected for classification.

*ReliefF* weighs features by taking into account their interactions. It uses the K-nearest neighbour method to weigh-up features which discriminate best from the neighbours of the different class. Thus, this method considers not only how strongly features are related to the observed class but also how distant they are from the opposite class. We set K to 23 — half of the total sample size, and retained 10 most discriminative features. ReliefF was implemented using scikit-rebate [14].

*mRMR* (minimum redundancy maximum relevance) is another filter method which selects features which have the highest relevance and at the same time lowest redundancy. It selects features which discriminate categories well but are dissimilar to each other. Both minimum redundancy and maximum relevance criteria are based on mutual information. Ten top ranked features were retained.

### E. Result validation

In this study nested CV was used [15]. Nested CV similarly to commonly used K-fold CV approach validates the results iteratively in CV folds, using all of the available data for training and also reusing all of it for testing. Both validation methods thus are economical and well suited when available data is small as is the case in this study. Nested CV is, however, different from K-fold CV in a significant aspect — it avoids pooling train and test data. When nested CV is performed a portion of data is split at the beginning of each CV fold for testing and a model is then developed on the reduced training set, including data normalisation feature selection and parameter tuning. This is repeated iteratively with splitting a different portion of

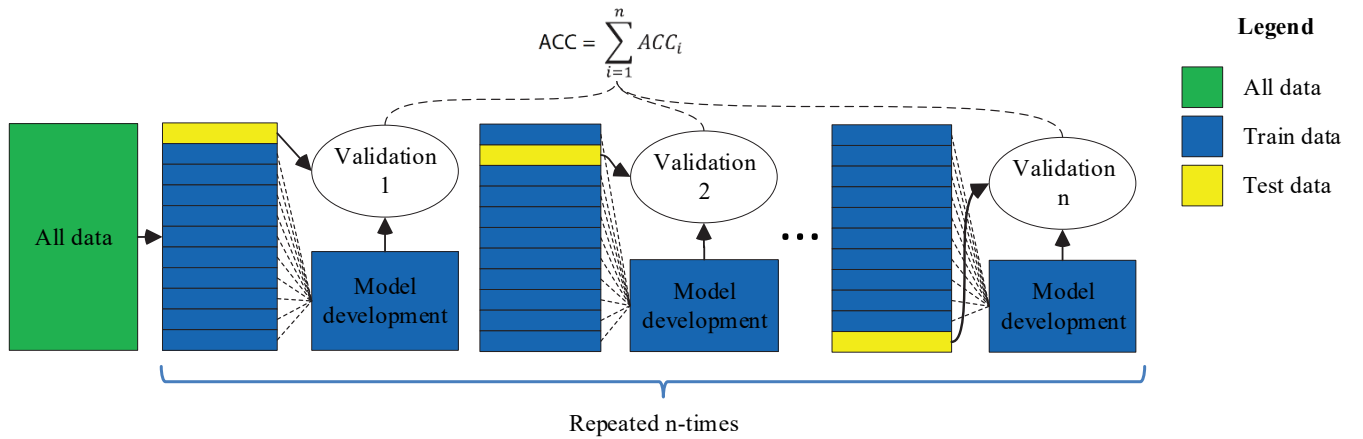


Fig. 2. Nested validation, ACC - overall accuracy of the model,  $ACC_i$  - accuracy in a single CV fold

the data for validation, and each time developing a new model for training from scratch until all of the data is used (Fig. 2). By using the nested CV approach test data is separate from model development and in that respect this approach is similar to Train/Test Split validation. Varma and Simon (2006) [8] have demonstrated that nested CV produces almost unbiased performance estimates, while K-fold CV approach, which pools train and test data, can produce significantly overoptimistic results. In this study 10 fold Nested CV was used and performance of the model was calculated as a mean performance of ten CV folds. Nested CV was performed 100 times by randomly splitting the data to training and testing sets to obtain performance distributions.

#### F. Result significance

Result significance was assessed with permutation testing. The labels of the data samples were randomly permuted 1000 times and empirical  $p$ -statistic calculated as in [16].

### III. RESULTS

#### A. Behavioural results

Here we include results only for measures which showed most prominent differences between groups (two-sample  $t$ -test, Table II). ASC participants performed pointing movements slower than TD participants as demonstrated by mean movement duration and velocity. ASC participants reached peak vertical amplitude earlier in the movement and peak velocity later in the movement compared to TD participants (Fig. 3). ASC participants, however, performed pointing movements more accurately than TD participants as indicated by the average area covered by separate points to a single target sticker location (Fig. 1a, Table II). These differences were significant and had medium to large effect sizes as indicated by Cohen's  $d$ .

#### B. Classification results

Results of SVM-RBF coupled with different feature selection methods showed that statistically significant classification accuracy of 71%, with a sensitivity of 75%, specificity of

TABLE II  
POINTING MOVEMENT AND POINTING ACCURACY RESULTS

	Duration (s)	Mean velocity (mm/s)	% time of vertical amplitude	Time to peak velocity	Overall area (mm <sup>2</sup> )
ASC mean	0.42	775.98	0.43	0.19	30.87
TD mean	0.37	911.96	0.47	0.17	64.01
ASC SD	0.08	142.89	0.05	0.03	16.15
TD SD	0.07	212	0.06	0.03	65.75
$t$ -statistic	2.47	2.57	2.77	2.34	2.29
$p$ -statistic	0.02	0.01	0.01	0.02	0.03
Cohen's $d$	0.74	0.77	0.82	0.69	0.81

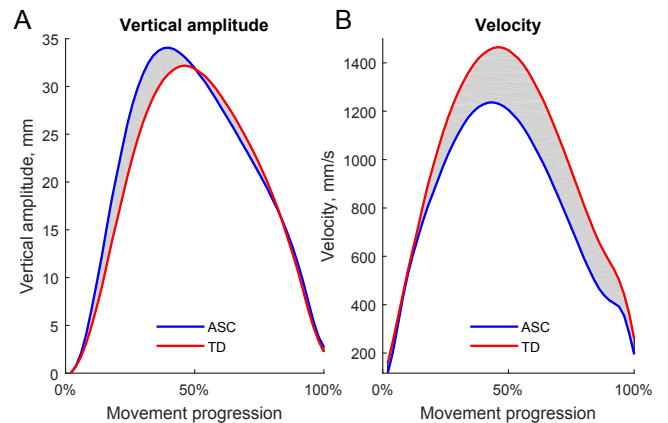


Fig. 3. Movement vertical amplitude (a) and velocity (b) averaged for ASC and TD participants. Shaded areas show the difference between groups.

TABLE III  
CLASSIFICATION RESULTS

Algorithm	Accuracy	Sensitivity	Specificity	$p$ -value
$t$ -test	71%	75%	66%	0.021*
ReliefF	70%	72%	68%	0.027*
SVM-RFE	62%	65%	59%	0.113
mRMR	60%	64%	55%	0.148

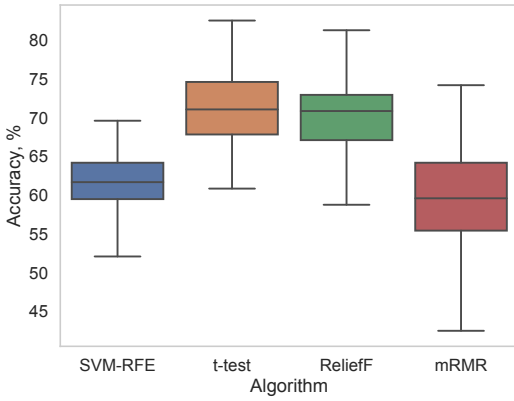


Fig. 4. Accuracy distributions for four models with different feature selection.

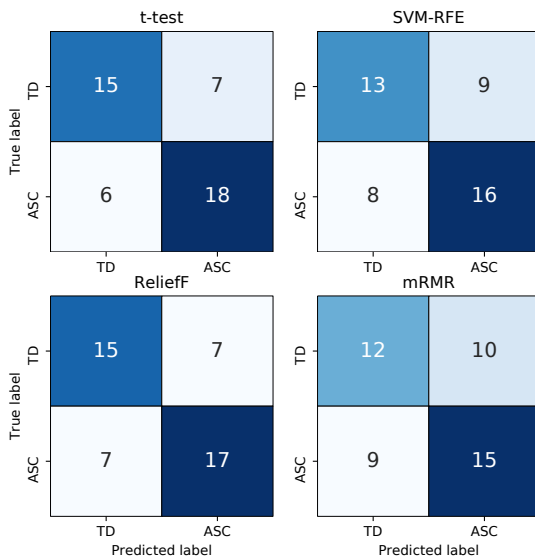


Fig. 5. Confusion matrices for four models with different feature selection.

66% and  $p = 0.021$  was achieved with  $t$ -test feature selection and 70% accuracy, 72% sensitivity, 68% specificity and  $p = 0.027$  with ReliefF feature selection. When SVM-RFE and mRMR feature selectors were used classification results were not significant (Table III, Fig. 4, Fig. 5).

#### IV. CONCLUSIONS

In this study, we explored whether kinematic characteristics of a short and easy to perform movement task can discriminate between ASC and TD individuals. Behavioural results showed that ASC individuals performed movements slower but more accurately. Our developed ML algorithms were capable statistically significantly predict ASC diagnosis. SVM coupled with  $t$ -test feature selection produced classification accuracy of 71%, with a sensitivity of 75%, specificity of 66% and  $p = 0.021$ . The study also shows that even if traditional parametric statistics methods show differences between groups with large effect sizes this does not necessarily translate to high ML classification performance. Our future work is centred on using kinematic data

of more complex movement imitation tasks and preliminary results are promising. In addition, to ensure the results are robust we also plan to validate models with newly collected independent data set.

#### REFERENCES

- [1] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders: DSM-5®*. Washington, DC: American Psychiatric Pub, 2013.
- [2] K. A. Fournier, C. J. Hass, S. K. Naik, *et al.*, "Motor Coordination in Autism Spectrum Disorders: A Synthesis and Meta-Analysis," *Journal of autism and developmental disorders*, vol. 40, no. 10, pp. 1227–1240, 2010.
- [3] M. Mari, U. Castiello, D. Marks, *et al.*, "The reach-to-grasp movement in children with autism spectrum disorder," *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, vol. 358, no. 1430, pp. 393–403, 2003.
- [4] C. M. Freitag, C. Kleser, M. Schneider, *et al.*, "Quantitative assessment of neuromotor function in adolescents with high functioning autism and Asperger syndrome," *Journal of autism and developmental disorders*, vol. 37, no. 5, pp. 948–959, 2007.
- [5] A. Anzulewicz, K. Sobota, and J. T. Delafield-Butt, "Toward the Autism Motor Signature: Gesture patterns during smart tablet gameplay identify children with autism," *Scientific Reports*, vol. 6, no. April, pp. 1–13, 2016. DOI: 10.1038/srep31107.
- [6] A. Crippa, C. Salvatore, P. Perego, *et al.*, "Use of machine learning to identify children with autism and their motor abnormalities," *Journal of autism and developmental disorders*, vol. 45, no. 7, pp. 2146–2156, 2015.
- [7] B. Li, A. Sharma, J. Meng, *et al.*, "Applying machine learning to identify autistic adults using imitation: An exploratory study," *PloS one*, vol. 12, no. 8, e0182652, 2017.
- [8] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, pp. 1–8, 2006.
- [9] E. Combrisson and K. Jerbi, "Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy," *Journal of Neuroscience Methods*, vol. 250, no. February, pp. 126–136, 2015.
- [10] M. Van Selst and P. Jolicoeur, "A solution to the effect of sample size on outlier elimination," *The Quarterly Journal of Experimental Psychology Section A*, vol. 47, no. 3, pp. 631–650, 1994.
- [11] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144–152, 1992.
- [12] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [13] I. Guyon, J. Weston, S. Barnhill, *et al.*, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, pp. 389–422, 2002.
- [14] R. J. Urbanowicz, R. S. Olson, P. Schmitt, *et al.*, *Benchmarking relief-based feature selection methods*, arXiv e-print. <https://arxiv.org/abs/1711.08477>, 2017.
- [15] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society*, vol. 36, no. 2, pp. 111–147, 1974.
- [16] M. Ojala and G. C. Garriga, "Permutation tests for studying classifier performance," *Journal of Machine Learning Research*, vol. 11, pp. 1833–1863, 2010.