General comments:

This was an interesting and generally well-written manuscript. My reviewing role was exclusively with respect to the application of machine learning (ML). I encourage the authors to 1) include the results from the ML analysis in the main text; 2) more clearly describe the process for splitting the training/test data and evaluating model performance; and 3) mention the limitations of the ML approach in the discussion section.

More detailed comments are provided below:

1. The inclusion of Supplemental Figure 1 in the main manuscript would help the reader evaluate the effectiveness of the classification algorithm. Otherwise, consider adding accuracy values in-text to support statements such as "This analysis indeed revealed that antigravity muscles contained important information, allowing separating age-groups with some of the best success-rates." [lines 312-314]

2. The phrasing on lines 277-280 make it difficult to distinguish whether: a) the data was split into training and test-sets in the traditional ML manner (i.e., model construction/training is performed on cross-validated training dataset, and a separate test-set is withheld for model evaluation); or b) whether five-fold cross-validation was performed on the full dataset and 'training and test sets' refer to the four training folds and one validation fold for each subset of the data (i.e., no test set was withheld to assess the models performance and generalizability on unseen data; see below for visualization).

<u>a) Train + test split (80/20 split as example)</u>

Full Dataset
- Training Data (80% of Full Dataset)
  - Fold 1: Train (80%) | Validate (20%)
  - Fold 2: Train (80%) | Validate (20%)
  - Fold 3: Train (80%) | Validate (20%)
  - Fold 4: Train (80%) | Validate (20%)
  - Fold 5: Train (80%) | Validate (20%)
- Testing Data (20% of Full Dataset) - Used for final model evaluation

OR

<u>b) Cross-validation on full dataset</u>

Full Dataset
- Fold 1: Train (80%) | Validate (20%)
- Fold 2: Train (80%) | Validate (20%)
- Fold 3: Train (80%) | Validate (20%)
- Fold 4: Train (80%) | Validate (20%)
- Fold 5: Train (80%) | Validate (20%)

I have reviewed the methods papers cited in the manuscript (Chambellant et al., 2023; Thomas et al., 2023) and, in conjunction with the present paper, my impression is that this method employs cross-validation across the full dataset and that an independent test-set was not withheld. Correct

me if I am mistaken. Either way, please clarify in the manuscript whether the model was evaluated on an unseen test set.

3.If there is a rationale for not testing model performance on unseen data, please provide supporting citations in-text for readers unfamiliar with this approach to ML. Otherwise, please mention the limitations of this approach in the discussion (i.e., limited ability of the model to generalize to unseen data, possibility of data leakage, limited ability to assess model overfitting). Finally, consider rephrasing lines 276-277 “To ensure robustness and generalization of the results, we employed a five-fold cross-validation method” to address the lack of generalizability when testing on the validation set (c.f., the test set).