# Understanding the dynamics of para rowing through machine learning

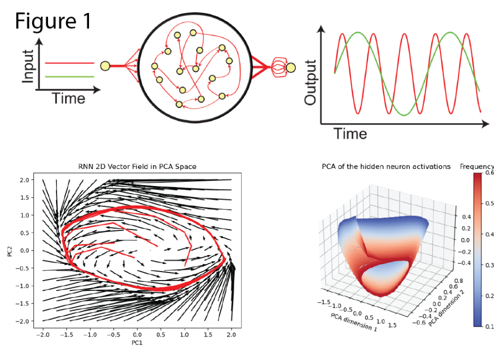
**Keith Murray (ktmurray@mit.edu), Emelie Eldracher (emelieel@mit.edu)**

## Abstract:

This investigation introduces a novel approach for understanding different Para-rower classification categories using single-camera video. We employ Recurrent Neural Networks (RNNs) to analyze the temporal dynamics of Para-rowing strokes. Through this, we aim to build a “rowing-stroke-generator” and shed light on the sparse field of Para rower’s biomechanics. This approach begins by processing 2D video data with MeTRAbs 3D Autoencoder to extract 3D coordinates [5]. By utilizing frame rate, we find the x, y, and z velocity, acceleration, and angles of each of 30 keypoints between frames. We introduce a 2-layer Dense Classifier, revealing the importance of keypoints on Para classification. Subsequently, we analyze an RNN trained on the keypoints’ time-series data to unveil patterns in the principal component analysis (PCA) representation. We also inspect the equivariance of our RNN, demonstrating insights into the relationship between motion-type, Para classification, and keypoint trajectories over time.

## Introduction:

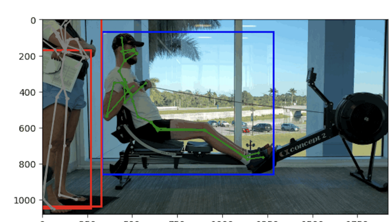
The two pre-existing artificial intelligence (AI) rowing tools are not optimized for Para athletes, and coaches have questioned the black-box methods behind their hidden models [4][2]. Three Para classifications exist for the varying disabilities of athletes: PR1, PR2, and PR3. PR1 involves arms-only rowing on a stationary seat, PR2 uses the upper body on a stationary seat, and PR3 includes leg movements with a sliding seat. While 17 research papers delve into Para rowing, only seven focus on biomechanics, only one uses AI via linear regression, and the largest data number of participants in any one study is 11 [3]. Our analysis explores this field at a level of machine learning unseen in prior work. By harnessing the power of RNNs, the time-series nature of a rowing stroke can be analyzed. We seek to train a recurrent neural network (RNN) to copy key points through time from Para-rowers. Our choice of RNN was motivated by the potential to create a ‘rowing-stroke-generator’ where interventions on rower’s biomechanics could be predicted. RNNs are also amenable to the ‘dynamics view’ and ‘the principle of equivariance.’

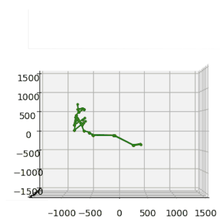



Previous work understanding the oscillatory dynamics of RNNs sought to train RNNs to be a sine wave generator [6]. Inputs to the RNN were time-invariant amplitudes ranging from 0.1 to 0.6 and the RNN was trained to produce sine waves with a frequency corresponding to the input. Principal component analysis (PCA) revealed that the trained RNN learned a low-dimensional cylinder, consisting of “stacked” limit-cycles, whereby the input would push the activity of the network to the limit-cycle on the cylinder corresponding to the desired output frequency (**Fig. 1**). For this work, we hypothesized that an RNN trained to output rowing key points would learn a similar low-dimensional cylinder where the length of the cylinder would represent PR class.

## Methods:

**Data Collection:** We gathered 6 videos featuring USRowing Para High Performance Athletes during the week of 11/27/23. Each Para Classification (PPR1, PR2, and PR3) was represented by two videos, each featuring a different athlete. The videos were all truncated to the same length of 328 frames.





Visual representation of MeTRAbs 3D plotting largest figure in the frame

**2D-3D:** We passed each frame of our video through MeTRAbs to extract 3D-point data. MeTRAbs was chosen for its superior performance to Strided Transformers and RIE [1]. The largest figure in the frame, chosen by bounding box size, was plotted and centered at the origin via their pelvis.

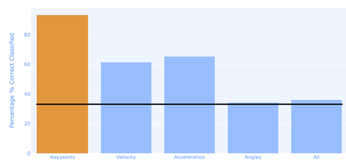
**Velocity, Acceleration, Joint Angle:** The largest person’s keypoint locations were then analyzed between frames to, using the frame rate, reveal velocity and acceleration of each keypoint. MeTRAbs utilizes a metric scale, so velocity and acceleration have units m/s and m/s^2. Angles between connecting joints were calculated by classical trigonometry.

## Modeling:

**Dense Classifier:** Using a 2-hidden layer deep dense classifier, we built a model to classify an athlete as PR1, P2, or PR3 using a singular frame. We performed an 80/20 train/test split on all frames from our 6 videos. Our input varied per training. Initially, we trained the model using only keypoint data for 24 joints. Subsequently, we separately input velocity, acceleration, angles between joints, and finally, all these parameters combined. This model used a learning rate of .001 with 10 epochs and a batch size of 32. We omitted 6 face-based keypoints as they were not relevant to this study.

**RNN training:** The particular RNN architectural variant we implemented was a gated recurrent unit (GRU) of 200 neurons. The input dimension was (328, 20) corresponding to a unique 20-dimensional random key vector that was assigned to each video in the dataset. The key vector was constant through time. The output dimension was (328, 72) corresponding to the (x,y,z) coordinates of 24 key points (6 key points were facial key points and were excluded) across 328 frames. The loss function was mean squared error and the optimization algorithm was AdamW with a learning rate of 1e-3 and a weight decay of 1e-2. Given the small number of videos in the dataset, we trained the model via curriculum training by introducing a new video to the training dataset every 500 epochs. There were 5000 training epochs in total and no testing dataset was used. All code was implemented using PyTorch.

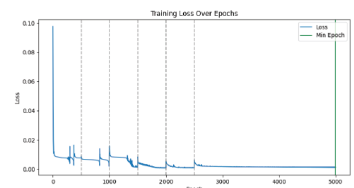
**Results:**




**Classifier Analysis:** The classifier underscored the significant role of keypoints in determining model accuracy. The x,y,z coordinates of athlete's keypoints alone substantially influenced inference accuracy. Although both velocity and acceleration data perform above chance in classification, we opted to prioritize keypoint data in our more complex temporal modeling.

**RNN Analysis:** We analyzed the RNN that achieved the lowest loss during training (**Fig. 2**). Through performing PCA on the rates of the RNN, we found that the RNN implemented a manifold of limit cycles (**Fig. 2**) in a manner similar to the sine wave generator RNN (**Fig. 1**). Interestingly, principal components (PC) 1 and 2 seemed to represent the periodicity of the key points generated while PC 3 seemed to represent the PR class of the video (**Fig. 2**). While PCA does not guarantee that all trained RNNs will use the same PC to represent task-relevant information, we found this particular PCA representation striking for the trained RNN.

Figure 2



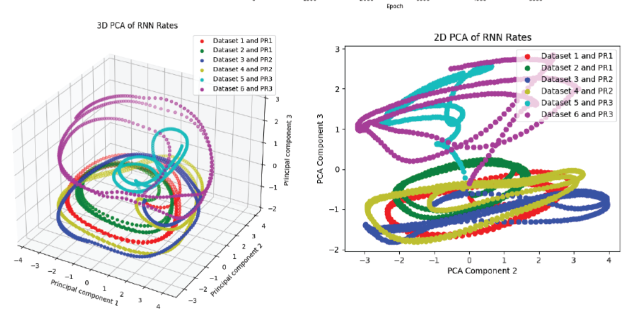
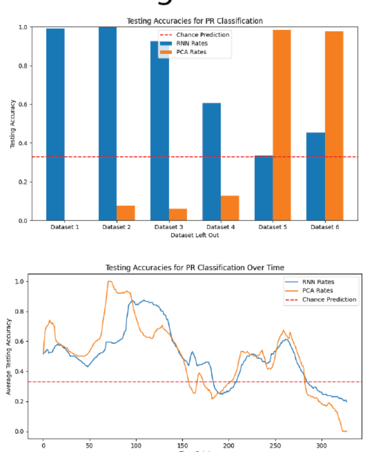




Figure 3




To numerically validate that PC 3 represents PR class, we trained a series of linear decoders to classify PR class from RNN and PCA rates. We implemented a multinomial linear regression model and cross-validated the testing accuracies by leaving one video out. Results in **Fig. 3** show that PR 3 videos are easily classified by PCA rates while PR 1 and PR 2 videos are easily classified by RNN rates. This shows that PCA 3 does represent the PR class, but only for PR 3 vs PR 1 and PR 2. It is peculiar that PR 1 and PR 2 classes are readily classified with RNN rates, but we did not investigate further.

To investigate the representation of the PR class in the trained RNN further, we sought to understand if this representation was equivariant with respect to time. The principle of equivariance can be mathematically represented as $g(f(x))=f(g(x))$. In the context of biomechanics and time, we formulated equivariance as $g(f(x,t))=f(g(x),t)$ where x represents the type of motion (i.e. rowing stroke, tennis serve, etc.), $g(x)$ represents the PR classification of x, and $f(x,t)$ represents the trajectory of key points through time. To quantify the extent to which the trained RNN exemplifies this principle, we trained a linear decoder at each frame of videos. We implemented a multinomial linear regression model and cross-validated the testing accuracies by leaving one video out. **Figure 3** shows the average testing accuracies through time in comparison with chance classification. Optimally, if the model were equivariant, then the classification accuracy through time would look like a horizontal line near 0.8 to 1.0 testing accuracy. Overall, the model displays some equivariance, but could become more optimal with a larger dataset.

**Conclusion:** In this investigation, we analyzed Para-rowing dynamics in a multidimensional AI-driven approach. By extracting 3D data from single-camera 2D video, we derive physics of the stroke: position, velocity, acceleration, and angles in metric space. Our findings indicate that keypoint analysis greatly influences a model's ability to classify a rower into PR1, PR2, or PR3. Our application of RNNs revealed a pattern of convergence between the PR1 and PR2 classes. This may provide commentary to a global debate on range of motion between PR1 and PR2 athletes. Moreover, our exploration of equivariance with respect to time suggests that, with more video data, clearer biomechanics patterns may emerge. Our study helps pave the way for more advanced AI application in Para Sport.

**References:**


[1] Baumgartner, T., Paassen, B. & Klatt, S. Extracting spatial knowledge from track and field broadcasts for monocular 3D human pose estimation. *Sci Rep* 13, 14031 (2023). https://doi.org/10.1038/s41598-023-41142-0

[2] Newby, J. (2022). WODProof Tackles Rowing Efficiency With Launch of AI Rowing Coach. https://morningchalkup.com/2022/09/01/wodproof-tackles-rowing-efficiency-with-launch-of-ai-rowing-coach/

[3] Puce, L., Biz, C., Trompetto, C., Marinelli, L., Currà, A., Cavaggioni, L., Formica, M., Vecchi, V., Cerchiaro, M. C., Trabelsi, K., Bragazzi, N. L., & Ruggieri, P. (2023). A Scoping Review with Bibliometric Analysis of Para-Rowing: State of the Art and Future Directions. *Healthcare (Basel, Switzerland)*, *11*(6), 849. https://doi.org/10.3390/healthcare11060849

[4] RowerUp. The first AI rowing technique video analysis platform for rowing on a boat and an erg. https://rowerup.com/faq

[5] Sárándi, I., Linder, T., Arras, K., & Leibe, B.. (2020). MeTRAbs: Metric-Scale Truncation-Robust Heatmaps for Absolute 3D Human Pose Estimation. arXiv:2212.14474. https://istvansarandi.com/dozens/

[6] Sussillo, D., & Barak, O. (2013). Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural computation*, 25(3), 626-649.