

Nov 13, 2024

# 🌐 Win Your Race Goal: A Generalized Approach to Prediction of Running Performance

DOI

[dx.doi.org/10.17504/protocols.io.ewov1d8d2vr2/v1](https://dx.doi.org/10.17504/protocols.io.ewov1d8d2vr2/v1)

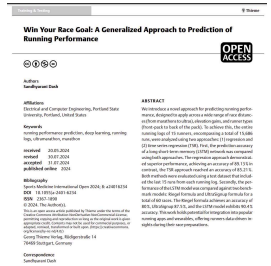
Sandy Dash<sup>1</sup>

<sup>1</sup>Portland State University



Sandy Dash

Portland State University



## Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.ewov1d8d2vr2/v1>

External link: <https://www.thieme-connect.de/products/ejournals/pdf/10.1055/a-2401-6234.pdf>,  
<https://github.com/SandyDash19/WinYourRaceGoal>

**Protocol Citation:** Sandy Dash 2024. Win Your Race Goal: A Generalized Approach to Prediction of Running Performance. protocols.io <https://dx.doi.org/10.17504/protocols.io.ewov1d8d2vr2/v1>

**Manuscript citation:**

S. Dash, "Win Your Race Goal: A Generalized Approach to Prediction of Running Performance," Sports Med. Int. Open, vol. 8, p. a24016234, Oct. 2024, doi: 10.1055/a-2401-6234.

**License:** This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Protocol status:** Working

**Please email any questions to** [dashs@px.edu](mailto:dashs@px.edu)

**Created:** November 12, 2024

**Last Modified:** November 13, 2024

**Protocol Integer ID:** 111945

**Keywords:** deep learning; marathon; running logs; running performance prediction; ultramarathon., predicting running performance, running performance, marathons to ultra, marathon, runner, performance of the lstm model, runner type, race distance, wide range of race distance, race goal, run, race preparation, lstm, lstm model, prediction accuracy, performance, last 15 run, race, benchmark model, popular running app, superior performance, accuracy, prediction, term memory, generalized approach to prediction

## Abstract

We introduce a novel approach for predicting running performance, designed to apply across a wide range of race distances (from marathons to ultras), elevation gains, and runner types (front-pack to back of the pack). To achieve this, the entire running logs of 15 runners, encompassing a total of 15,686 runs, were analyzed using two approaches: (1) regression and (2) time series regression (TSR). First, the prediction accuracy of a long short-term memory (LSTM) network was compared using both approaches. The regression approach demonstrated superior performance, achieving an accuracy of 89.13% in contrast, the TSR approach reached an accuracy of 85.21%. Both methods were evaluated using a test dataset that included the last 15 runs from each running log. Secondly, the performance of the LSTM model was compared against two benchmark models: Riegel formula and UltraSignup formula for a total of 60 races. The Riegel formula achieves an accuracy of 80%, UltraSignup 87.5%, and the LSTM model exhibits 90.4% accuracy. This work holds potential for integration into popular running apps and wearables, offering runners data-driven insights during their race preparations.

## Attachments



[WinYourRaceGoal](#)

971KB

## Guidelines

The running logs collected for this study has been in adherence to IRB Protocol #207107–18. The dataset uploaded on GitHub is completely anonymized.

## Troubleshooting

Published article: <https://www.thieme-connect.com/products/ejournals/pdf/10.1055/a-2401-6234.pdf>

1

Code: <https://github.com/SandyDash19/WinYourRaceGoal>

2

**Can there be a unified mechanism to predict a runner's performance in a race of any distance or elevation gain?**

- 3 One way of achieving the above goal is by using runner's entire running log and an autoregressive deep learning model known as Long Short Term Memory (LSTM).
- 3.1 Data preparation which involves data clean up, taking care of missing data and performing seasonal decomposition to get an idea of the complexity of the model required to make accurate prediction.
- 3.2 LSTM model with an architecture elaborated in the published article and the github repo.
- 3.3 Use data augmentation techniques if you are working with a small dataset. Sliding window technique is used in published article.
- 3.4 Determine what regularization techniques you will use to prevent overtraining. For e.g. Early Stopping algorithm is used in the published article. Patience which is a hyperparameter that determines the number of times the increase in validation loss is tolerated before terminating epoch loop.
- 3.5 To increase the robustness of the hyperparameters perform k-fold cross validation during hyperparameter search.
- 3.6 The idea is that by learning the running log of a runner the LSTM captures both long-term and short-term correlations among exogenous and endogenous variables, even without explicitly incorporating the notion of time into the network.



- 3.7 It is demonstrated in the journal article that with enough running history the LSTM model is able to generalize very well and predict with an average of 90.4% accuracy even for unseen races up to two years in the future. To compare the LSTM model's performance against two benchmark models; Riegel formula and UltraSignup formula, 60 races of varying distances described in table 2 in the article were chosen. 54% of these races occurred after the data collection cut off i.e. December 2021.