

TACTICAL ANALYSIS USING MACHINE LEARNING

Troesch S.¹, Rossi M.² and Wehrlein J.P.¹

¹Swiss Federal Institute of Sport, Section for Elite Sports, Magglingen, Switzerland

²Swiss Triathlon Federation, Ittigen, Switzerland

KEYWORDS: Machine learning, tactical analysis, triathlon mixed team relay, Tokyo 2020

INTRODUCTION

The **triathlon mixed team relay**, in which four athletes successively complete a short triathlon (figure 1) will be new to the Olympic program in 2020.

To optimize race and team selection tactics, the **relative importance of different race sections** (e.g. run of fourth athlete) must be understood.

Machine learning algorithms can quantify the importance of different race sections for the prediction of the final relay ranking.

RESEARCH QUESTION

How important are the different sections of the triathlon mixed team relay for the prediction of the final result?

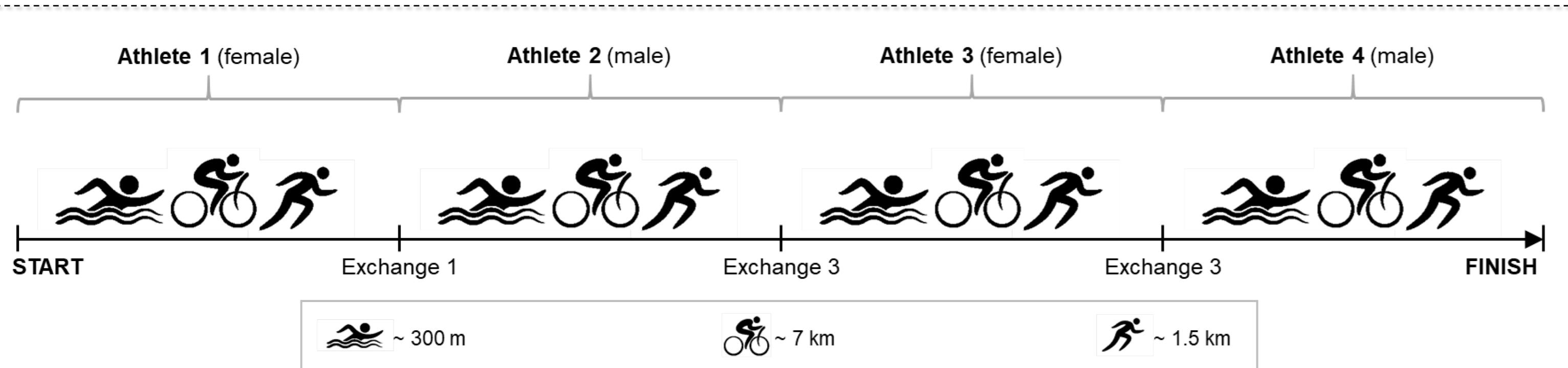


Figure 1: Race structure of the triathlon mixed team relay.

METHODS

- A “random forest” model¹ was fitted to the results of the **12 major relay races** (major championships, world triathlon series) of the 2018 season.
- Input variables were **swim, bike and run times** of all four legs (i.e. athlete 1 to athlete 4). The independent variable was the final team ranking.
- Variable importance scores** were derived via the decline in model accuracy (MSE) with random variable permutation².

RESULTS

The importance scores of all race sections are shown in Figure 2.

The **highest importance scores** were found in the **run sections** of athlete 3 (importance score = 4.2) and athlete 1 (importance score = 2.9).

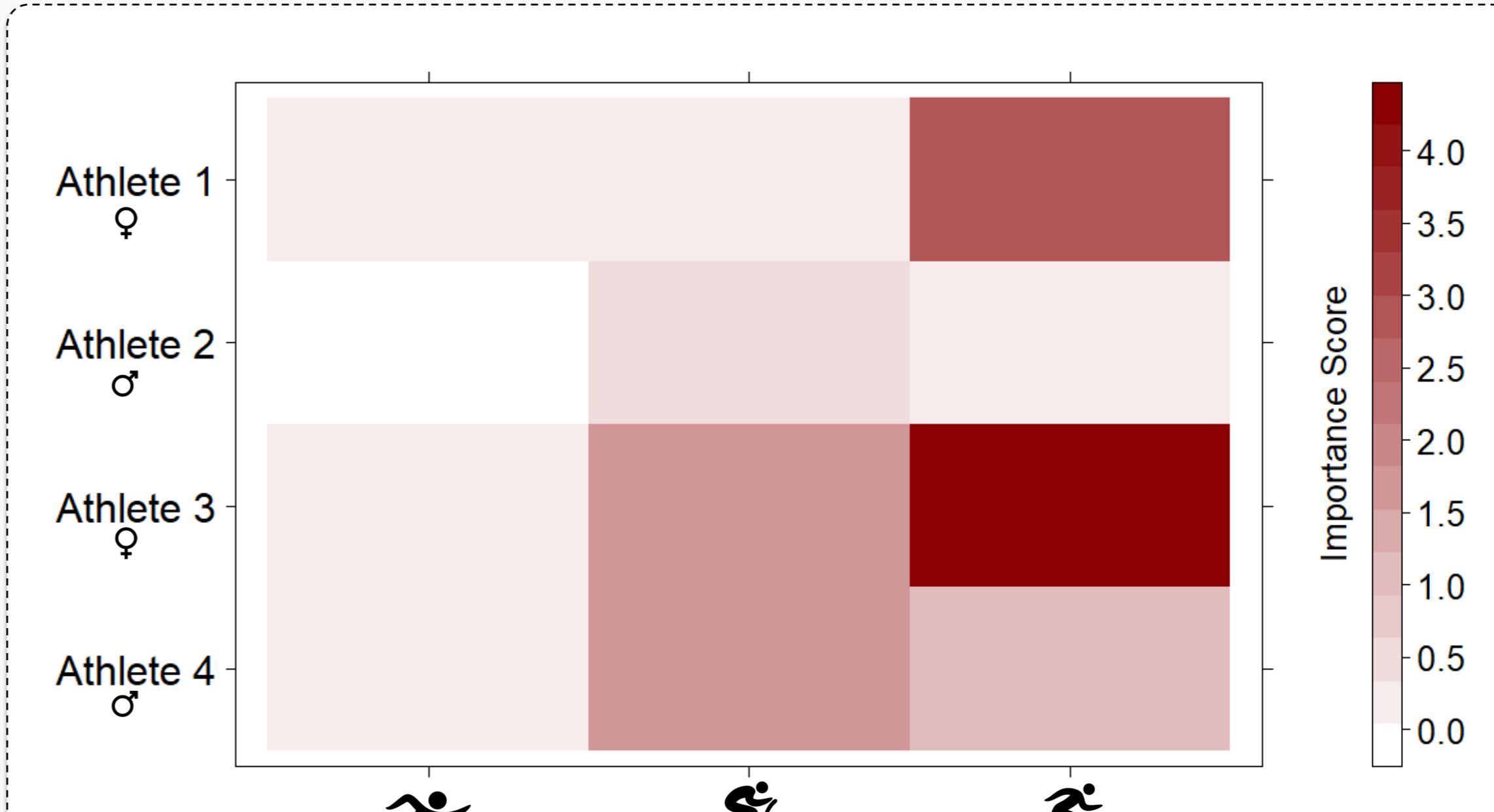


Figure 2: Importance of different race sections (e.g. run of third athlete) for the prediction of the final relay ranking. Darker red = higher importance = larger impact on final ranking.

DISCUSSION

The analysis revealed that the most important sections for the prediction of the final relay ranking are run and bike of athletes 3 and 4 as well as the run of athlete 1. That is, fast times particularly in these sections are connected to a good final result of the team.

This knowledge can be used by coaches to select the relay lineup or to optimize the training regimen of athletes.

SO WHAT!?

- The study contributed to a **better understanding of race tactics** in the triathlon relay.
- Also in **every other sport**, machine learning models can assess the contribution of sub-results to final competition rankings.
- Like this, a **better understanding of any discipline** can be built.

¹ Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

² Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. BMC bioinformatics, 9(1), 307.