Title: Predicting marathonian performances in sport medicine from deep neural networks

Partners : IBISC (univ Evry, université Paris-Saclay)
Basic AI and Data Science : statistical training in big dimensions
Specialized ML and AI : signal, image, vision
Application domain : precision medicine, modern sport medicine
Mots-clés deep learning, machine learning, deep tech, biosignals, modern sport practice, precision medicine
Total duration of internship: 6 months (graduate) or 3 months (undergraduate)
Working period: From 2024/02/01 to 2024/09/01

Context and objectives

We propose to develop an analysis of the physiological parameters of marathon runners, based on deep neural networks to deduce physiological conclusions on their performances [1].

During evolution, human physiology was optimized to cover great distances to find enough food to support the brain's metabolism. The popularity of the marathon among humans is a legacy of our ability to run long distances using aerobic metabolism. The number of participants in the London Marathon has fallen from 7,000 to 35,000 over the past 30 years and participation in road races in general has increased by more than 50% over the past decade. This popularity is characterized by the emergence of amateur marathoners. They now have GPS heart rate meters and seek to run in a speed or heart rate (HR) zone without theoretical bases and experience a drastic drop in their speed from the 25th km. Other runners prefer to run "by feel" and a significant challenge is to analyze and interpret their physiological responses (signals) to help them improve their performance. For example, comparing the analysis of heart rate (HR) in the first half and last quarter of the half marathon, when the runner started to have glycogen depletion (a major cause of the decrease in speed at the end of the race) highlights a change in the dynamics of the FC.

To understand how the performance of marathon runners is induced through endurance training as a function of exercise intensity and duration, we analyze the physiological modulation of these runners based on their physiological constants. A set of sensors ECG, O2 saturation, etc. record physiological parameters throughout the race. The variability of activity between runners will be studied to define how intense and intermittent efforts, distances covered, rest times, etc. are linked to performance. Ultimately, it’s about finding how the marathon runner’s performance can be optimized during a race.

Are the constants the same during the race? Are the physiological modulations of the same order in marathon runners? Can we capture these modulations with our sensors? Can we find a non-linear statistical model of the deep neural network type?

No study to date has investigated how to predict the physiological effort of the marathon runner or how to “rank” a runner based on their physiological parameters, rather than their performance.

Methodology

Deep learning analysis of multi-sensor data in sports is uncommon. The reason is the lack of performance data, often acquired on a small number of athletes due to their cost of acquisition and the variability of cohorts, etc. The difficulty of obtaining a sufficient amount of specific reliable training data for supervised machine learning requires the investigation of new strategies. A solution suggested by very recent studies [3], proposes

to develop new generic functions such as fractal methods or to use the data augmentation method to generate signals. Learning by neural networks inspired by ladder networks or regime or auto-encoding networks. Thus, based on the information shared at this stage, two algorithmic approaches seem interesting to us:

- use deep learning variational autoencoder,

- adapt modern image recognition neural networks dedicated to edge computing known to be light and not very CPU intensive.

Numerous publications in various disciplinary fields (physics, geology, medical imaging, etc.) have already demonstrated the effectiveness of multimodality for machine learning.

Expected results
The expected solution will make it possible to better characterize the changes in the dynamics of the marathon runner's race by associating multiple weak signals with physiological constants. It will improve the reading quality of sports performances because they take too long to detect.
This solution will be able to determine which information in the signals implies which performance. AI can help the sports doctor by determining in advance which performance tests to evaluate first. AI can make this initial assessment. Ultimately, the technique should provide more information than the human eye.

Expected performance criteria:
The evaluation of the new procedure compared to a referenced procedure raises numerous methodological difficulties. The expected performance indicators are
1. the repeatability of the performance characterization process in a degraded situation or not, in the case of lack of data or a reduced performance test.
2. the effectiveness of the tool to be tested in ground truth and quantified
3. a speed of execution of a few minutes.
Expected results

References
[1] Guillaume Saës, Wejdene Ben Nasr, Stéphane Jaffard, Florent Palacin, Véronique Billat. Analyse Multifractale des données physiologiques de marathoniens. *GRETSI 2022, XXVIIIème Colloque Francophone de Traitement du Signal et des Images*, Sep 2022, Nancy, France. ⟨hal-03694475v2⟩
[2] A.L. Goldberger. Is the normal heartbeat chaotic or homeostatic ?. News Physiol. Sci., 1991
[3] E. Wesfreid, V. Billat et Y. Meyer. Multifractal analysis of heartbeat time series in human races. *Appl. Comput. Harmon. Anal.*, 2005
[4] V. Billat, F. Palacin, M. Correa et J.-R. Pycke. Pacing strategy affects the sub-elite marathoner’s cardiac drift and performance. Frontiers in Psychology, 2020.

Profile and skills required
The recruited person will be in the 3rd year of engineering school or Master's. He will be able to understand and develop adaptive learning algorithms and to process medical dataset, index it and use it in an operational system to achieve the mission described above.
Programming skills: Python or C / C ++. A practice of Tensorflow and Pytorch would be a plus. The practice of French is not compulsory. His(her) English is fluent. The work will be carried out at the IBISC Laboratory located on the Evry campus of the UPSaclay. IBISC develops multidisciplinary, theoretical and applied research in the field of information

sciences and engineering, with a strong orientation towards health applications. The selected candidate will have the chance to work in an interdisciplinary team and with a consortium of data scientists and clinicians from the CHSF. The project is multidisciplinary, at the interface of machine learning, computer science and medicine.

Supervision and scientific and material conditions
The student will be supervised by Vincent Vigneron, Jean-Philippe Congé and Hichem Maaref from the IBISC laboratory (Univ d'Évry, Université Paris-Saclay). All master machine learning, signal and image processing.

Contact: please send marks (Bachelor and Master) to

Vincent Vigneron, Veronique Billat and Hichem Maaref
{hichem.maaref,vincent.vigneron}@univ-evry.fr Phone: +33 6 63 568 760
veronique.billat@billatraining.com