

Titre: Predicting marathonian performance in sport medicine from deep neural networks

Partners : IBISC (univ Évry, université Paris-Saclay)

Basic AI and Data Science : apprentissage statistique en grandes dimensions

Specialized ML and AI : signal, image, vision

Application domain : médecine de précision, modern sport

Mots-clés deep learning, machine learning, deep tech, biosignals, modern sport practice, precision medicine

Laboratoires partenaires impliqués : IBISC (UEVE) **durée totale du stage** 6 mois (M2) ou 3 mois (M1) **date de début et de fin du stage** 15/02/2023 au 1/09/2023

Nous proposons de développer une analyse des paramètres physiologiques des marathonien, basée sur des réseaux de neurones profonds basés dans le but d'en déduire des conclusions physiologiques sur leurs performances [1].

Au cours de l'évolution, la physiologie humaine a été optimisée pour couvrir de grandes distances afin de trouver suffisamment de nourriture pour soutenir le métabolisme du cerveau. La popularité du marathon chez les humains est un héritage de notre capacité à courir de longues distances en utilisant le métabolisme aérobie. Le nombre de participants au marathon de Londres est passé de 7 000 à 35 000 au cours des 30 dernières années et la participation aux courses sur route en général a augmenté de plus de 50% au cours de la dernière décennie. Cette popularité est caractérisée par l'émergence de marathonien amateurs. Ils disposent désormais de cardiofréquence mètres GPS et cherchent à courir dans une zone de vitesse ou de fréquence cardiaque (FC) sans bases théoriques et subissent une chute drastique de leur vitesse à partir du 25ème km. D'autres coureurs préfèrent courir "à la sensation" et un défi important est d'analyser et d'interpréter les séries temporelles de leurs réponses physiologiques pour les aider à améliorer leurs performances. Par exemple, la comparaison de l'analyse de la fréquence cardiaque (FC) dans la première moitié et le dernier quart du semi-marathon, lorsque le coureur a commencé à avoir un épuisement du glycogène (cause majeure de la diminution de la vitesse à la fin de la course) met en évidence un changement dans la dynamique de la FC.

Objectives

Pour comprendre comment les performances des coureurs de marathon sont induites grâce à l'entraînement en endurance en fonction de l'intensité et de la durée de l'exercice, nous analysons la modulation physiologique de ces coureurs en fonction de leurs constantes physiologiques. Un ensemble de capteurs ECG, saturation en O₂, etc.

enregistrent pendant toute la course les paramètres physiologiques. La variabilité de l'activité entre les coureurs sera étudiée pour définir comment les efforts intenses et intermittents, les distances parcourues, les temps de repos, etc. sont reliés aux performances. In fine, il s'agit de trouver comment les performances du marathonien peuvent être optimisées pendant une course.

Les constantes sont-elles identiques pendant la course ? Les modulations physiologiques sont-elles du même ordre chez les marathonien ? Peut on capturer ces modulations avec nos capteurs ? Peut on trouver un modèle statistique non linéaire de type deep neural network.

Aucune étude à ce jour n'a étudié comment prédire l'effort physiologique du marathonien ou comment « ranker » un coureur selon ses paramètres physiologiques, plutôt que sur ses performances.

Methodology

L'analyse par apprentissage profond des données multicateurs en sport est peu fréquente. La raison est le manque de données de performances, souvent acquises sur un petit nombre de sportifs en raison de leur coût d'acquisition et de la variabilité des cohortes, etc. La difficulté d'obtenir une quantité suffisante de données d'entraînement fiables spécifiques pour une apprentissage automatique supervisée nécessite l'étude de nouvelles stratégies. Une solution suggérée par des études très récentes [3], propose de développer de nouvelles fonctions génériques comme les méthodes fractales ou d'utiliser la méthode d'augmentation de données pour générer des signaux. L'apprentissage par réseaux de neurones inspirés des réseaux en échelle ou des réseaux de régime ou d'auto-encodage. Ainsi, sur la base des informations partagées à ce stade, deux approches algorithmiques nous semblent intéressantes : - utiliser le deep learning variational autoencoder, - adapter les réseaux de neurones modernes de reconnaissance d'images dédiés au edge computing connus pour être légers et peu gourmands en CPU.

De nombreuses publications dans divers domaines disciplinaires (physique, géologie, imagerie médicale, etc.) ont déjà démontré l'efficacité de la multimodalité pour l'apprentissage automatique.

Expected results

La solution attendue permettra de mieux caractériser les changements de la dynamique de la course du marathonien en associant de multiples signaux faibles aux constantes physiologiques. Il améliorera la qualité de lecture des performances sportives car trop longues à détecter. Cette solution sera en mesure de déterminer quelles informations dans les signaux impliquent quels performances. L'IA peut aider le médecin du sport en déterminant à l'avance les tests de performances à évaluer en premier. L'IA peut faire cette évaluation initiale. À terme, la technique devrait donner plus d'informations que l'œil humain.

Critères de performance attendus :

L'évaluation de la nouvelle procédure par rapport à une procédure référencée soulève de nombreuses difficultés méthodologiques. Les indicateurs de performance attendus sont 1. la répétabilité du processus de caractérisation des performances en situation dégradée ou non, dans le cas de manque de données ou d'un test de performance réduit. 2. l'efficacité de l'outil à tester en vérité terrain et à quantifier 3. une rapidité d'exécution de quelques minutes.

References

[1] Guillaume Saës, Wejdene Ben Nasr, Stéphane Jaffard, Florent Palacin, Véronique Billat. Analyse Multifractale des données physiologiques de marathonien. *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 candidate must provide a full CV mentioning his age, his notes since the 12th year of Bachelor as well as a cover letter.

The recruited person will be in the 3rd year of engineering school or Master's. It 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 Évry 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.

Encadrement et conditions scientifiques

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

Contact: Vincent Vigneron, Hichem Maaref

{hichem.maaref,[vincent.vigneron](mailto:vincent.vigneron@univ-evry.fr)}@univ-evry.fr Phone: +33 1 69 47 75 45, and Véronique Billat, veronique.billat@billatraining.com