Multi-class computer-aided diagnosis for prostate cancer mapping

Internship proposal

Scientific context Computer aided detection (CAD) systems are designed to assist clinicians in their diagnosis by highlighting abnormal regions in an image. One way of doing so consists in performing a classification at the voxel level where the decision model is learned on a series of feature vectors selected from normal and pathological locations on patient scans. This internship focuses on developing a new computer-aided diagnosis system for prostate cancer screening based on multiparametric magnetic resonance imaging. This CAD system will address a complex and challenging problem of prostate cancer mapping where the goal is to predict not only a presence/absence of cancer but also the degree of its aggressiveness. This work builds on previous works performed in the team to prototype a CAD system that performs binary classification based on multiparametric MRI [1, 2, 3]. From the methodological point of view, we plan to approach it by introducing new machine learning algorithms that tackle the problem induced by the presence of highly correlated and interdependent outcomes in multi-class classification. A main assumption of this project is that multi-class classification can be efficiently reduced to a set of binary classification problems for each possible class and further considered as a set of tasks in the multitask learning[4] where the interdependency and high classes confusability are appropriately addressed through adding new appropriate terms to the optimized objective function. The expected benefits of this idea lie in the proper mathematical modelling of the above mentioned phenomena and in the natural capacity of multitask learning to: (1) model the task relatedness that will allow to benefit from information provided by highly correlated and potentially interdependent classes and (2) operate on partially labeled data.

Objectives The primary goal of this internship is to perform an empirical study regarding the usefulness of multitask learning in handling classification tasks with multiple classes that can be interconnected. This study will be performed on the real-world data set consisting of a collection of annotated MRI images of prostates and will be further followed by an extension of the currently available multi-target algorithms to new previously unused loss functions [5]. These loss functions will rely on the concepts from the metric learning domain and will allow to account for important geometrical properties of the medical data.

Keywords Prostate cancer mapping, machine learning, biomedical imaging, multi-task learning, multi-class classification.

Skills Strong knowledge in at least one of the following fields is required:

- Image processing;
- Statistical learning (machine learning);
- Applied mathematics;
- Statistics.

The available code is written in Matlab and Python but other languages can be used. The successful candidate is expected to be autonomous and show strong motivation and interest in multidisciplinary research (image processing and machine learning in a medical context).

Supervision and collaborations The candidate will be supervised by Ievgen Redko and Carole Lartizien from CREATIS who have an acknowledged experience with multi-modal medical image processing and have developed strong skills in machine learning for medical imaging aver the past few years. This project will be part of the recently funded ANR project PERFUSE and will be carried out in collaboration with Massih Reza-Amini from LIG laboratory who will share his expertise in multitask learning in general and its application to multiclass classification in particular.

Applications Interested applicants are required to send a cover letter, CV and any other relevant documents (reference letter, recent transcripts of marks,...) to:
`carole.lartizien@creatis.insa-lyon.fr` and `ievgen.redko@creatis.insa-lyon.fr`.

Gratuity ≈ 550 €/mois.

References

[1] Rahaf Aljundi, Jérôme Lehaire, Fabrice Prost-Boucle, Olivier Rouvière, and Carole **Lartizien**. Transfer learning for prostate cancer mapping based on multicentric MR imaging databases. In *Machine Learning Meets Medical Imaging - First International Workshop, MLMMI 2015, Held in Conjunction with ICML 2015, Lille, France, July 11, 2015. Revised Selected Papers*, pages 74–82, 2015.

[2] Emilie Niaf, Rémi Flamary, Olivier Rouvière, Carole **Lartizien**, and Stéphane Canu. Kernel-based learning from both qualitative and quantitative labels: Application to prostate cancer diagnosis based on multiparametric MR imaging. *IEEE Trans. Image Processing*, 23(3):979–991, 2014.

[3] Léo Gautheron, Ievgen **Redko**, and Carole **Lartizien**. Adaptation de domaine pour la détection automatique du cancer de la prostate en imagerie irm multiparamétrique. *GRETSI*, 2017.

[4] R. A. Caruana. Multitask learning: a knowledge-based source of inductive bias. In *Proceedings of ICML 1993*, pages 41–48, 1993.

[5] M. Balcan, A. Blum, and N. Srebro. Improved guarantees for learning via similarity functions. In *Proceedings of COLT 2008*, page 287–298, 2008.