A web-based framework for visualization, annotation, and automatic exploitation of high-resolution bioimages using tree-based machine learning methods

Benjamin Stévens[1,2,*], Loïc Rollus[1,2,*], Louis Wehenkel[1,2], and Raphaël Marée[1,2,3]

[1]GIGA-Systems Biology & Chemical Biology, GIGA-R, University of Liège, Belgium
[2]Systems and Modeling Research Unit, Montefiore Institute, University of Liège, Belgium
[3]GIGA Bioinformatics Core Facility, GIGA, University of Liège, Belgium
*These two authors contributed equally to this work.

Introduction

Human interpretration of large-scale bioimaging datasets is time-consuming and somehow subjective. This stresses the need for replicable and generic computational methods to automate the extraction of useful, quantitative, information from these images. Towards that goal, there is a strong need for intuitive and efficient tools to foster collaboration between life science researchers working with high-throughput and high-resolution imaging equipments (such as whole-slide virtual microscopy, high-content screening, etc.), and computer scientists who need large and realistic training datasets to develop effective machine learning and computer vision methods.

This poster presents the ongoing development of a web-based software platform for remote visualization, collaborative annotation, and automated analysis of high-resolution biological images. The potential of such a framework is illustrated in the context of ongoing studies within the CYTOMINE research project (2010-2013) about lung tissue recognition and quantification in whole-slide Hematoxylin&Eosin stained images of experimental mice, and cell counting in whole-slide images of LBA cytopathology, from the Laboratory of Tumor and Development Biology (Prof. D. Cataldo, GIGA-Cancer, ULg), and from the Department of Pathology at Erasme University Hospital (Prof. I. Salmon, ULB).

Visualization and annotation of high-resolution images on the web

We rely on Java/JavaScript/REST technologies and a spatial, relational database to allow collaborative annotations of images with user-defined ontology terms.
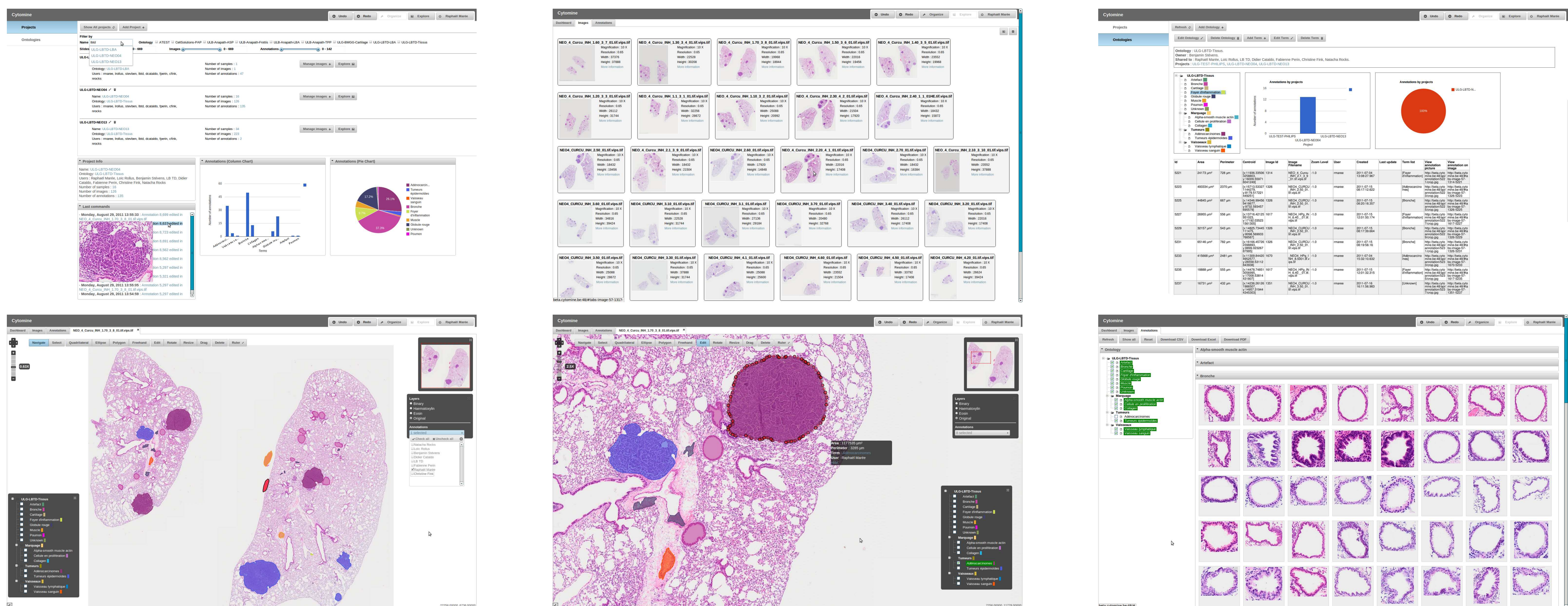
Integration of image processing and machine learning algorithms

Ongoing integration of generic machine learning algorithms using tree-based approaches [1,2,3] is expected to allow scientists to explore more efficiently their large-scale cellular and tissular data through ready-to-use content-based image retrieval, image segmentation and classification techniques exploiting training sets of manual annotations.
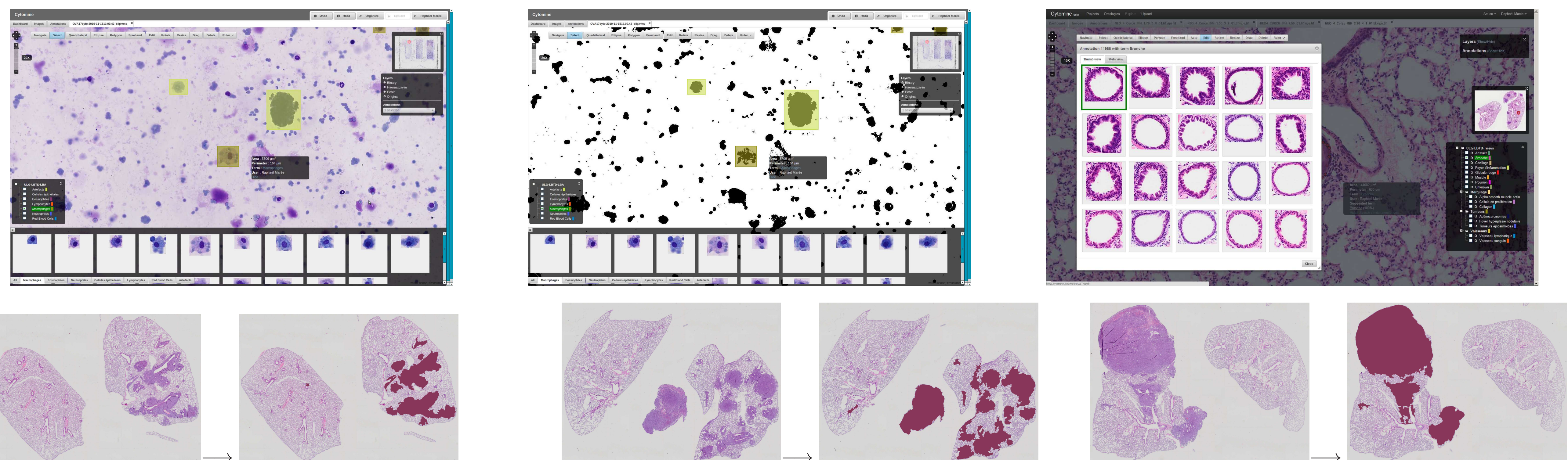
1 Marée et al., *Random Subwindows and Extremely Randomized Trees for Image Classification in Cell Biology*, BMC Cell Biology 2007.
2 Dumont et al., *Fast Multi-Class Image Annotation with Random Subwindows and Multiple Output Randomized Trees*, VISAPP 2009.
3 Marée et al., *Incremental Indexing and Distributed Image Search using Shared Randomized Vocabularies*, MIR 2010.

Acknowledgments