

# Semantic Integrative Digital Pathology. Insights on micro-semiological semantics and image analysis scalability

D. Racoceanu<sup>a,\*</sup>, F. Capron<sup>b</sup>

<sup>a</sup>*Sorbonne Universités, UPMC Univ Paris 06, CNRS, INSERM, Laboratoire d'Imagerie Biomédicale (LIB), F-75013, Paris, France*

<sup>b</sup>*Sorbonne Universités, UPMC Univ Paris 06, UH Pitié-Salpêtrière-CFx, Department of Pathology, APHP, UIMAP, F-75013, Paris, France*

---

## Abstract

Being able to provide traceable and dynamic second opinion has become an ethical priority for patients and healthcare professionals in modern computer aided medicine. In this perspective, a semantic cognitive virtual microscopy approach has been initiated<sup>1</sup>, by focusing on digital pathology. This approach supports the elaboration of pathology-compliant daily protocols dedicated to breast cancer grading, in particular mitotic count and nuclear atypia. A proof of concept has so been elaborated and an extension of these approaches is now on going under a collaborative digital pathology framework<sup>2</sup>.

Important milestone on the way to routine Digital Pathology, a series of pioneer international benchmarking initiatives have been launched for mitosis detection<sup>3</sup>, nuclear atypia grading<sup>4</sup> and glandular structures detection<sup>5</sup> some of the fundamental grading components in diagnosis / prognosis. These initiatives allow envisaging a consolidated validation referential-database for Digital Pathology in the very near future. This reference database will need coordinated efforts from all major teams working in this area worldwide, and will certainly represent a milestone for the acceptance of all future imaging modules in clinical routine.

In line with recent advances of molecular imaging and genetics, keeping the microscopic modality at the core of the future digital systems in pathology is fundamental to insure the acceptance of these new technologies, as for a deeper systemic, structured comprehension of the pathologies. After all, at the scale of routine Whole Slide Image (WSI) (around 0.22  $\mu\text{m}$  per pixel), the microscopical image represents a structured "genomic cluster", enabling

---

\*Corresponding author, E-mail: daniel.racoceanu@upmc.fr

<sup>1</sup>MICO project (COgnitive MIcroscopy) - French National Research Agency - Technologies for Health and Autonomy (ANR TecSan): <http://daniraco.free.fr/projects.htm>

<sup>2</sup>FlexMIm project (Collaborative Pathology) - Consolidated Interministerial Fund (FUI - Fonds Unique Interministériel) : <http://www.systematic-paris-region.org/en/projets/flexmim>

<sup>3</sup>Mitosis detection challenges: MITOS @ Int. Conf. Pattern Recognition (ICPR) Tsukuba, Japan, 2012: [http://ludo17.free.fr/mitos\\_2012/](http://ludo17.free.fr/mitos_2012/) and AMIDA @ Int. Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI) Osaka, Japan, 2013

<sup>4</sup>Mitosis detection and nuclear atypia grading challenge: MITOS&ATPIA @ Int. Conf. Pattern Recognition (ICPR) Stockholm, Sweden, 2014: <http://mitos-atypia-14.grand-challenge.org/>

<sup>5</sup>Glandular structures detection challenge: GlaS @ Int. Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI) Munich, Germany, 2015

a naturally structured integration of these heterogeneous data, a natural structured support for Integrative Digital Pathology (IDP) approaches.

In order to support this integration, as to structure the integration these heterogeneous information, a major effort is and will continue to be devoted to morphological micro-semiology (microscopical morphology semantics). Beside supporting the traceability of the results and the orchestration of high-content image analysis modules, the role of semantics will be crucial in the correlation between digital pathology and non-invasive medical imaging modalities, as in defining models able to make the link between traditional microscopy and recent label free technologies.

The challenge of massive visual data represents an obvious characteristic of Digital Pathology. Designing an operational integrative microscopy framework, needs to focus on scalable multi-scale imaging formalism. In this sense, we prospectively consider some of the most recent scalable methodologies adapted in Digital Pathology as Marked Point Process for nuclear atypia and Point-set Mathematical Morphology for architecture grading. Orchestrator of this scalable framework, the semantic-based WSI management (analysis, exploration, indexing, retrieval, report generation support) represents an important mean towards integrative big biomedical data approaches.

This insight reflects our vision, through an instantiation of essential bricks of this type of architecture. The generic approach introduced here is applicable for number of challenges related to molecular imaging, high content image management as, more generally, to bioinformatics.

*Keywords:* digital pathology, integrative digital pathology, semantics, virtual microscopy, big data, cancer grading, breast cancer, high-content image exploration

---

## 1. Introduction

Pathology diagnosis is a major element of severe and chronic disease. Despite the integration of digital modalities in teaching and research, their daily clinical use is still to be done. Future anatomopathological services need to use these digital technologies in valid routine pathological diagnosis and healthcare protocols, by integrating the Whole Slide Images (WSI) observation for diagnosis purposes in a whole large specific Digital Pathology case record (including clinical, radiologic and biological data). The goal is to describe, conceive and formalise an integrative framework of all these data, most of them already used - in different traditional formats - for the final diagnosis and prognosis. This will generate an Integrative Digital Pathology (IDP) process in which the innovation relies in linking the microscopic exam of WSI (standard stains, immunohistochemistry, hybridisation), to specific or generic annotations defined as micro-semiology semantic references. This enables the generation of a structured and standardised image-related report. Therefore, pathological data would be deposited together with those provided by the clinic, the radiology and the clinical biology (big data). IDP is supposed to act as a specialised and standardised patient file generator, to be linked - at the hospital level - with the patient medical record.

IDP generation of initiatives need to create a dynamic link between micro-semiology repository (data warehousing) and routine-annotated multi-scale images, in an intrinsic context of big data (high-content imaging) in histopathology. The role of the micro-semiology repository (ontology for anatomopathology - mainly related to the microscopic

scale of the exam) is to guide the orchestration of the WSI analysis, in order to consolidate / insure its reliability and traceability. This will be able to generate an traceable second opinion, based on specific quantitative support. Precise quantification support needs to be provided by highly scalable and semantic-driven imaging modules, able to generate a valuable second opinion, validated on a referential annotated database.

All the procedures, specifications and designs need to be generic, allowing a scalable extension to other diseases, as to other hospitals. Therefore, a coordinated standardisation effort is a key factor to facilitate this process.

Besides, a sustainable policy as a security-compliance need to be insured by actively involving the hospital information system.

Ethics and quality insurance need also to be strictly considered during the management of semantic catalogues, pathology reports, data-warehouse, as WSI cohort.

## 2. Histopathology Examination

Histopathology examination represents a milestone of the diagnostic and therapeutic decision support in severe chronic or acute diseases, malformative, inflammatory, metabolic and neoplastic. Concretised by the pathology report, histopathology examination it is an important part of the patient chart and requisite to all therapeutic decisions in oncology, for example during the MultiDisciplinary Team (MDT) meeting. It relies on individual medical professional diagnostic judgment, integrating gross and microscopic morphological criteria issued from standard and special techniques such as immunohistochemistry, enzymology, and hybridisation in situ. It has an integrative value, consolidated by clinical, radiological and biological contexts - among which, genetic and molecular.

The future of histopathology is obviously Digital (data and images). The challenge is to conciliate, in the framework of the healthcare, various usual missions as: performing the diagnosis of the case in the present moment, warehousing the medical data for the patient record, and also feeding and structuring the research strategy - particularly in oncology.

The present challenge we all need to face is to create an innovative healthcare pathway monitoring for which the milestones and the impact are related to: the production of digital histopathology tools, the modelling of the pathway and the conceptualisation of the associated massive database for high performance integrative cognitive research. The goal is to create generic standard and adapted software deployable in other fields of medicine for which bed to repository data collection, modelling of practice and knowledge are traced in situ.

IDP is capable to overcome the present limitations by merging pathology divers actors and tools, in order to defeat challenges as: the contractualized report-time delivery, the complexity of the procedures, the utilisation of in situ molecular techniques, the pharmacodiagnostic characteristics of the lesion/tumour according to the needs of targeted and personalised therapy.

Beyond a simple replacement of the microscopy by a screen or a tactile interface, this corresponds essentially to the integration of WSI, Regions of Interest (ROI), annotations, codes, and contextual graphs, in a real-time report building approach. This innovative pathology tracing observation and labelling need to become the safeguard of quality, security, and validity of diagnosis and data collection. This can be expected feasible for biopsy

interpretation mandatory for the first step diagnosis in a precise and standardised health care pathway.

What is on the way to be designed for a theme (i.e. senological<sup>6</sup>) shall be extended as a generic method to other pathways, including clinical, radiological, histopathological, biological, molecular signature, MDT and personalised therapy, in different phases of the cancer care (i.e. liver and prostate cancers).

### 3. Evolution of Digital Pathology

Digital approaches in Pathology are far more challenging than for Radiology (for example) for some particular reasons, related to high-content imaging acquisition, storage, analysis, complexity, heterogeneity, visual exploration and annotation. Besides, highly visual-memory work of pathologist requests years of practice before reaching the expertise. A huge amount of complex visual features need to be correlated with knowledge in the mind of the Pathologist before taking the responsibility of the medical report. The pressure of increasing number of pathological exams vs. relative decreasing number of pathologists urges us all to support this profession with advanced semantic modelling and visual analysis tools able to tackle the challenge of big data.

A single WSI is generally about 140,000 by 60,000 pixels RGB (3 bytes per pixel), meaning about 8 gigapixels. Each image is usually about 1 to 2GB when compressed, and 15 to 25GB when uncompressed. At the Pathology Department of the Pitié-Salpêtrière hospital, about 1500-2000 slides / day need to be analysed and diagnosed. For example, in APHP<sup>7</sup> we count about 30 anatomo-pathological centres, meaning about 300 petabytes of data per year.

In the next generation of projects leaded here, we plan to focus on a selected and specific range of specimens: core biopsies that are performed in a well defined healthcare protocol. A massive immediate switch to total digital for all types of requests is not yet realistic. However, for some major fields of medicine, integrating and modelling clinical and morphological data represents, in our opinion, a wise strategy.

The evolution of digital technologies together with the emergence of big-data technologies, allow us to realistically head for full digital in pathology in the next few years. The elaboration of these technologies in open-source formalisms need to generate a reference in Digital Pathology.

Starting by analysing, conceptualising and standardising the knowledge representation using a formalism adopted by the pathologists, we are now studying appropriate knowledge management, ontologies alignment and consolidation technologies, as high-content biomedical images exploration and annotation, all dedicated to big data warehousing in healthcare (including security, privacy and anonymisation constraints etc.). Generating Pathology reports from these semantic annotations will allow us to cross an important milestone in Digital Pathology. This will enable the pathologists to push the granularity or the semantic representation to the level of the micro-semiology in the whole medical chain.

---

<sup>6</sup>In the case of Pitié-Salpêtrière Hospital<sup>0</sup>, Paris, France, IDP initiative applies in priority to senology, due to strategic and motivating impact on all local actors.

<sup>7</sup>APHP: Public Hospital System of Paris and its suburbs (Assistance Publique des Hôpitaux de Paris)

Therefore - for example - the RadLex lexicon used in Radiology, mostly topological, could be extended with much finer granularity.

A dynamic operational framework is requested by creating a dynamic link between micro-semiology repository (at the data warehousing level) and routine-annotated multi-scale images, in an intrinsic context of big data (high-content imaging) in histopathology. Operational quantification support is to be provided by highly scalable and semantic-driven imaging modules, able to provide a valuable second opinion validated on a referential annotated database. All the imaging modules need to act in a routine semantic-guided pathway, validated by Pathologists. Finally, the designed system needs to be able to facilitate the readings, annotation and report generation in routine. This allows designing and building the necessary tools to work on massive WSI cohorts, in clinical routine.

The ANR TecSan MICO<sup>8</sup> project allowed Pathology Department in GHPS to elaborate the first proofs of concept of WSI exploration protocols for mitosis detection and nuclear atypia. The contextual modelling of the approach has also been realised [1], as the first proof of concept of semantic-driven imaging [2].

Further on, we need to built real-scale demonstrator for semantic knowledge management visual exploration support algorithms, enabling us to design, produce, test and validate all the necessary tools for routine Digital Pathology. This need to be done with an active support of all departments adjacent to the Pathology one (i.e. Clinical unit - Senology, Medical Imaging, Medical Biology and Medical Informatics). Therefore, the algorithms and the approaches need to become scalable to big visual data, a major characteristic of Digital Pathology. These initiatives are designed to produce management tools for modern massive WSI cohorts, semantically consistent and annotated according to an ACP-compliant protocol.

Knowledge representation analysis, conceptualisation and standardisation as active dynamic framework of semantic imaging are the core of future Digital Pathology platforms. Major actions need to be launched around: contextual modelling of standardised medical workflow, ontology alignment and semantic data-warehouse repository, scalable semantic-driven algorithms for massive high-content WSI analysis and process analysis of patient, information, material, and the financial analysis of the medical flows impacted.

This insight is focusing on scalable semantic-driven algorithms for massive high-content WSI analysis, by showing the principle of the semantic framework and the major categories of proposed scalable methods.

#### **4. Scalable quantification methods for Digital Pathology**

Generally speaking, considering the needs for scalability in Digital Pathology, we could mention three essential viewpoints:

- Mining high-content, large scale images in more expressive ways than a purely statistical description
- Developing frameworks allowing incorporating high-level geometric and shape priors for modelling the data
- Creating a new generation of image analysis, requiring highly automated algorithms being robustly scalable

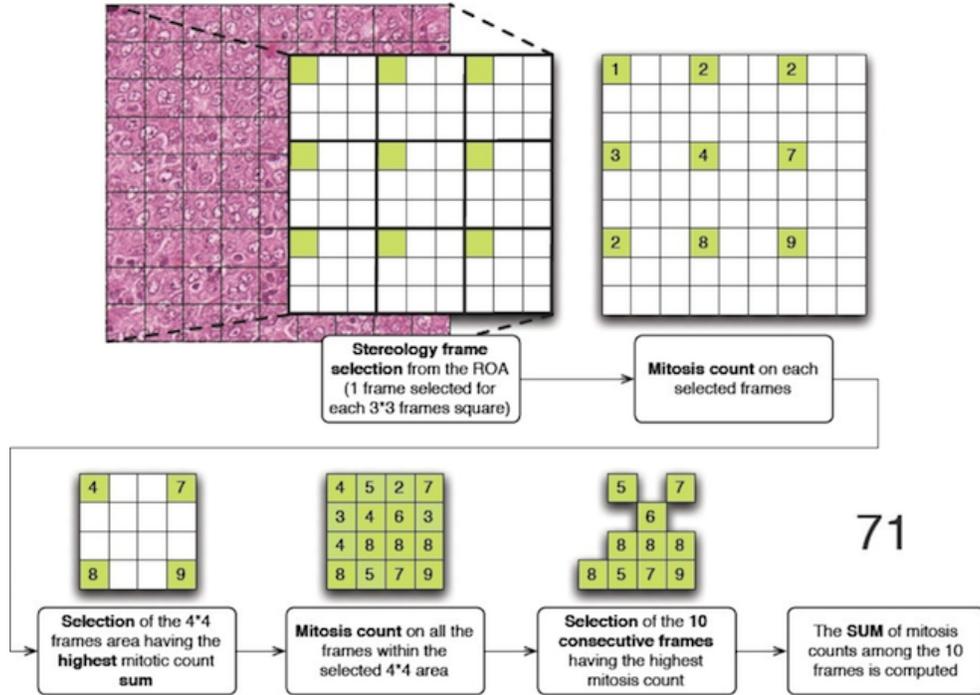


Figure 1: Example of mitotic count digital protocol [2]

In order to ensure the technology acceptability, Digital Pathology need to be able to provide a series of form extraction capabilities, compliant with pathology routine and semantics. The microsemiology to be used in image analysis and quantification support need to correspond to a sustainable ontology dedicated to microscopical features relevant for the pathology report.

We structure our ideas around the major imaging challenges, associated to three categories of challenges in digital pathology, each of them corresponding to a range of magnification: nuclei detection and segmentation, regional nuclei assessment and architecture analysis.

#### 4.1. Nuclei Detection and Segmentation

An important state of the art of the methods for nuclei detection, segmentation, and classification in Digital Pathology has been published in [3]. Usually, the analysis at the nuclei level request working with images at 40X magnification.

Mitotic score is one of the three Nottingham criteria. This score is computed from the region with the highest mitosis density within the WSI. The digital strategy proposed in [2] for the highest mitosis density evaluation without exhaustive analysis of the WSI is represented in Fig. 1.

The corresponding digital protocol starts by extracting the territory corresponding to the invasive area to be graded. This relevant territory (usually extracted by the Pathologist) is split into 40X frames. Since we are looking for the area having the highest concentration of mitotic nuclei, in order to accelerate the critical ROI search, we use a sampling on the set of frames covering the tumour (see Fig. 1) (see [2]).

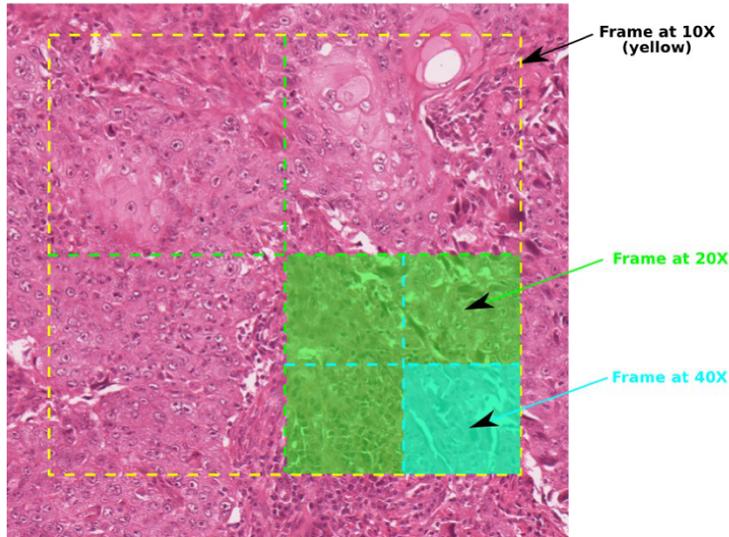


Figure 2: WSI exploration model for digital pathology [2].

This protocol provides a clinically acceptable mitosis score approximation. However, more details are needed for its realisation, like the different steps of the strategy, and the type of data exchanged from one step to another. Therefore, mitotic scoring flow has to be split in modular algorithms, allowing a precise description in terms of inputs and outputs, in a more accurate manner. Accordingly, several modular algorithms are designed from the mitotic count strategy as a first step towards a cognitive platform. In order to insure a broad usability of these modules, a particular attention is paid to the definition of the algorithms interfaces [2].

Significant series of initiatives have been recently launched by structuring a mitosis database with a proper annotation and a precise diagnosis purpose. MITOS<sup>9</sup> @ ICPR 2012 has been the first such initiative [4, 5], followed by AMIDA<sup>10</sup> @ MICCAI 2013 and refined by MITOS&ATYPIA<sup>4</sup> @ ICPR 2014. The goal is to build-up a validation database dedicated (in this case) to mitotic count for breast cancer. Similar initiatives shall be come in the near future for different pathologies.

#### 4.2. Regional Nuclei Assessment

The first challenge concerning the nuclear atypia has been organised ICPR 2014<sup>4</sup>. The annotation has been done at two magnification levels: the operational one at 20X (mostly used for nuclear atypia assesment), with the usual grades used in Pathology (1 to 3) as with a detailed description of relevant nuclear atypia patterns (size of nuclei, size of nucleoli, density of chromatin, thickness of nuclear membrane, regularity of nuclear contour and anisonucleosis (i.e. size variation within a population of nuclei) at each 40X frames inside the 20X one, according the Digital Pathology protocol set up (see Fig. 2, [2]).

<sup>9</sup>Mitosis detection challenge: MITOS @ Int. Conf. Pattern Recognition (ICPR) Tsukuba, Japan, 2012

<sup>10</sup>Mitosis detection challenge: AMIDA @ Int. Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI) Osaka, Japan, 2013

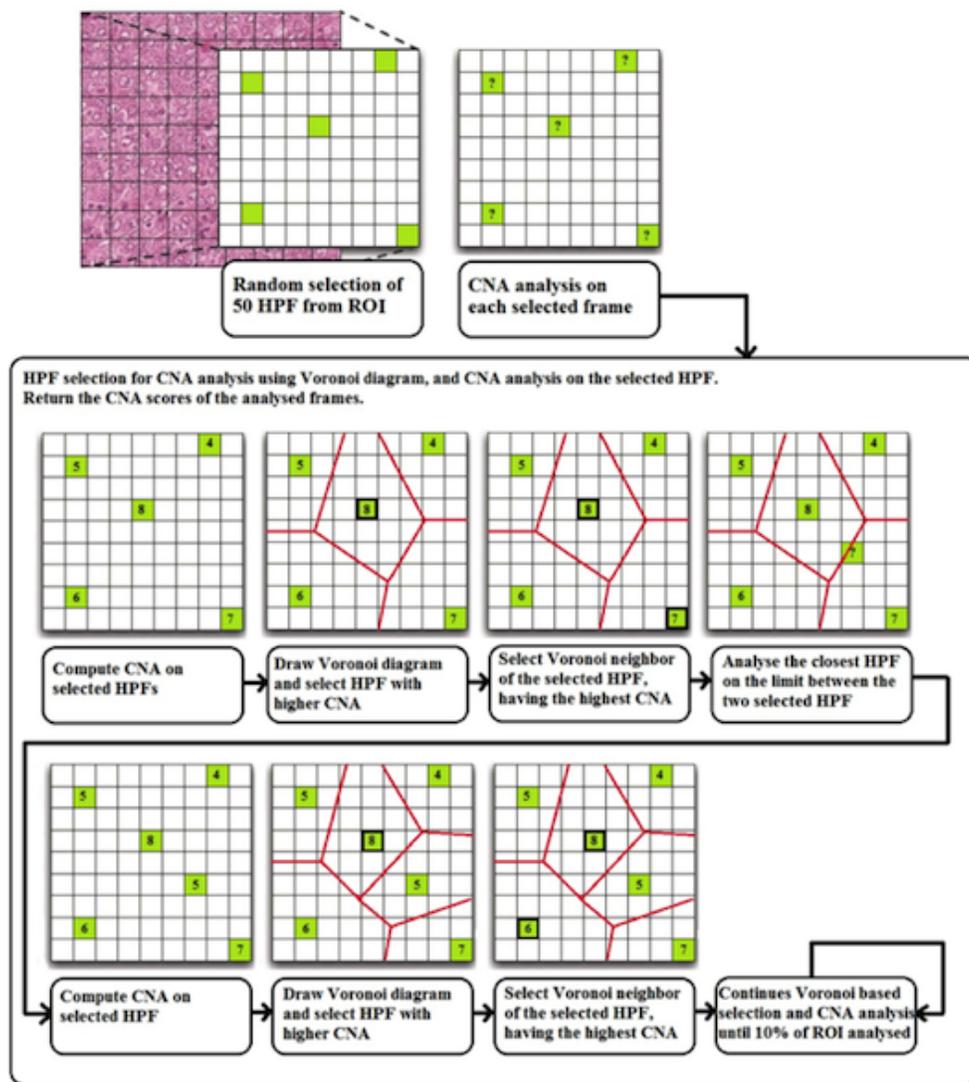


Figure 3: Example of nuclear atypia grading digital protocol [6]

Concerning the nuclear atypia grading protocol, one has been proposed inspired by the use of dynamic Voronoi diagrams [6] (see Fig. 3):

An interesting scalable approach of nuclei detection nuclear atypia assessment has been published in [7]. Combined with a probabilistic filtering based on scale, shape, and colour, this has been shown to give interesting results in H&E images [8]. The parallelisation of this method opens interesting perspectives to WSI analysis in the near future [9, 10].

#### 4.3. Architecture Analysis

The study of nuclei's architecture is of a major interest for a series of pathologies. This is usually done at 5X to 10X magnification. In breast cancer, the tubular structures (one of the three H&E grading criteria) corresponds to such example, in which nuclei arrangement need to be analysed with different imaging tools (i.e. graphs). A recent challenge has been launched (Glas<sup>5</sup> @ MICCAI 2015) concerns the detection of glands in colon cancer.

A good example of scalable method based on mathematic morphology and graphs [11] in

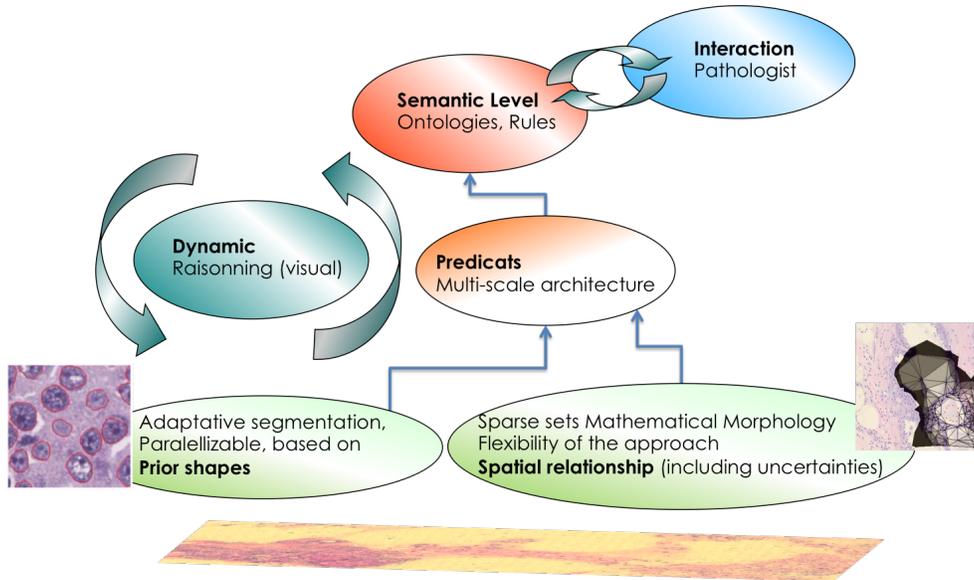


Figure 4: Overall integration of radiometric and architectural scalable imaging functions

inflammatory Bowel Disease is presented in [12], a study done in the frame of the FlexMIm<sup>2</sup> project. Indeed, developing a mathematical morphology based on graphs allows to rely on a sparse structure of epithelial nuclei (for example), in order to launch spatial relationship operators (surrounded by, between, inside, liking ...). An interesting study of the ontologies extension able to take into account these relationships has been done in [13]. This work need to be extended and deployed further on, in order to deal with a series of very useful such spatial descriptors for diagnosis and prognosis support.

## 5. Semantic-Driven Whole Slide Imaging Analysis in Digital Pathology

Basically, histopathology is a highly visual-cognitive medical discipline. Instead of building a hard-coded imaging protocol, we choose to use a semantic framework [2] enabling us to deploy a traceable, flexible approach for a semantic WSI exploration and digital pathology assistance. Based on a dedicated maintainable ontology and a reasoner this semantic platform aims at supporting pathologists during the diagnosis process, by taking benefit of updated medical and image processing technologies, while relying on valid histopathology experts' knowledge.

In order to extend the methodology described in [2] and effectively orchestrate an important number of imaging tools by the knowledge and the reasoning, this system will need to use scalable imaging methods, able to deal with big data challenges triggered by routine digital pathology and huge imaging data. Such scalable imaging tools need to be adapted to the needs of digital pathology according to two family of methods detailed in the previous sections: statistical and graph-based (see Fig. 4).

Such semantic imaging architecture offers benefits as traceability of the decision support, improved technology acceptance, increased flexibility of the use of WSI processing algorithms by choosing the priority of the application (quality, efficiency, compliance etc.).

## 6. Conclusion

This insight paper illustrates our vision concerning a semantic approach of Integrative Digital Pathology, able to deal with traceable, modulable and scalable imaging algorithms for high-content image analysis. The objective is to make histopathological WSI analysis reliable and traceable, for an improved second opinion, based on quantitative support.

Indeed, semantic orchestration allows working with high-content imaging web-services, by creating, therefore, a way to stimulate the whole community in this area and helping them to solve the Intellectual Property issue related to image analysis algorithms.

Creating a professionally annotated database on each digital pathology challenge will constitute a validation referential for the future of Integrative Digital Pathology. All the imaging tools will need to pass this validation test before being considered for routine protocols.

Besides, scalability and parallelisation capabilities are essential nowadays in order to go for massive high-content challenges in an operational context.

Our future initiatives will go along this vision, together with our medical, industrial and academic partners involved in our national and european projects. Through this insight paper, we hope increasing this community, by crystallising the premises of a future standard for semantic-driven and scalable high-content imaging framework.

Corresponding to an ambitious challenge, leading to a potential breakthrough in integrative digital pathology, this work has also the meaning of its genericity. Indeed, the approach is applicable for number of additional challenges, related, for example, to cytology, biology, molecular imaging and, in general, to a large set of high-content high-throughput automatic screening approaches.

## References

- [1] E. Attieh, P. Brézillon, and F. Capron, "Context-based modeling of an anatomo-cytopathology department workflow for quality control," *Proceedings of Modeling and Using Context (CONTEXT 2013)*, vol. 8175, pp. 235–247, 2013.
- [2] D. Racoceanu and F. Capron, "Towards semantic-driven high-content image analysis. an operational instantiation for mitosis detection in digital histopathology," *Computerized Medical Imaging and Graphics*, vol. 2, pp. 2–15, June 2015.
- [3] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A review – current status and future potential," *IEEE Reviews on Biomedical Engineering*, no. 7, pp. 97–114, 2014.
- [4] L. Roux, D. Racoceanu, N. Loménie, M. Kulikova, H. Irshad, J. Klossa, F. Capron, C. Genestie, G. Le Naour, and M. Gurcan, "Mitosis detection in breast cancer histological images: An icpr 2012 contest," *Journal of Pathology Informatics*, vol. 4, May 2013.
- [5] L. Roux, D. Racoceanu, N. Loménie, M. Kulikova, H. Irshad, J. Klossa, F. Capron, C. Genestie, G. Le Naour, and M. Gurcan, "Mitosis detection in breast cancer histological images," in *ICPR International Contest*, (Tsukuba, Japan), November 2012.

- [6] A. Veillard, N. Loménie, and D. Racoceanu, “An exploration scheme for large images: application to breast cancer grading,” in *Proceedings of the 20th international conference on pattern recognition*, 2010.
- [7] M. Kulikova, A. Veillard, L. Roux, and D. Racoceanu, “Nuclei extraction from histopathological images using a marked point process approach,” in *SPIE Medical Imaging*, (San Diego, California, USA), February 2012.
- [8] A. Veillard, M. Kulikova, and D. Racoceanu, “Cell nuclei extraction from breast cancer histopathology images using color, texture, scale and shape information,” in *11th European Congress on Telepathology and 5th International Congress on Virtual Microscopy*, (Venice, Italy), June 2012.
- [9] C. Avenel and M. Kulikova, “Marked point processes with simple and complex shape objects for cell nuclei extraction from breast cancer H&E images,” in *SPIE Medical Imaging*, (Orlando, USA), February 2013.
- [10] C. Avenel, P. Fortin, and D. Béréziat, “Parallel birth and death process for cell nuclei extraction in histopathology images,” in *International Conference on Parallel Processing*, (Lyon, France), Octobre 2013.
- [11] N. Loménie and D. Racoceanu, “Point set morphological filtering and semantic spatial configuration modeling: Application to microscopic image and bio-structure analysis,” *Pattern Recognition*, vol. 45, no. 8, pp. 2894–2911, 2012.
- [12] B. Ben Cheikh, P. Berthaud, and D. Racoceanu, “Preliminary approach for crypt detection in inflammatory bowel disease,” *RITS*, 2015.
- [13] A. Tutac (ép. Branici), *Formal Representation and Reasoning for Microscopic Medical Image-Based Prognosis. Application to Breast Cancer Grading*. PhD thesis, International co-tutelle - Politehnica University of Timisoara, Romania and Université de Franche Comté, Besançon, France, 2010.