Querying medical imaging datasets using spatial logics (position paper)⋆

Gina Belmonte[1], Giovanna Broccia[2], Laura Bussi[2,3], Vincenzo Ciancia[2], Diego Latella[2], Mieke Massink[2]

[1] Azienda Toscana Nord Ovest, S. C. Fisica Sanitaria Nord – Lucca, Italy
[2] Consiglio Nazionale delle Ricerche, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" – Pisa, Italy
[3] Università di Pisa, Dipartimento di Informatica – Pisa, Italy

Abstract. Nowadays a plethora of health data is available for clinical and research usage. Such existing datasets can be augmented through artificial-intelligence-based methods by automatic, personalised annotations and recommendations. This huge amount of data lends itself to new usage scenarios outside the boundaries where it was created; just to give some examples: to aggregate data sources in order to make research work more relevant; to incorporate a diversity of datasets in training of Machine Learning algorithms; to support expert decisions in telemedicine. In such a context, there is a growing need for a paradigm shift towards means to interrogate medical databases in a semantically meaningful way, fulfilling privacy and legal requirements, and transparently with respect to ethical concerns. In the specific domain of Medical Imaging, we envisage the adoption of query languages that can unambiguously express semantically rich queries on possibly multi-dimensional images, in a human-readable, expert-friendly and concise way. Our approach is based on querying images using Topological Spatial Logics, building upon a novel spatial model checker called VoxLogicA, to execute such queries in a fully automated way.

Keywords: Open health data platforms, Spatial Logics, Model Checking

1 Introduction and related work

A number of technologies with proven disruptive impact in Computer Science have revolved around domain-specific data models and query languages. Let us just name a few. The *Structured Query Language SQL* [40] revolutionised data representation and access, and is nowadays one of the pillars of modern

Information Technology. The *eXtensible Markup Language XML* [41] and its query/transformation languages *XQuery, XPath, XPointer*, and siblings, provide solid grounds to any modern data-centric or document-centrinc infrastructure. The javascript library *jQuery*[4] is the de-facto standard for traversing HTML documents and identifying elements to be transformed using javascript. These technologies (and many others) have in common the adoption of a concise, unambiguous, declarative *query language* that domain experts, without any particular computer programming skill, can use, enabling widespread adoption of a number of transformative key functionalities. It is not an overstatement to say that without the invention of such fundamental technologies, most of the modern applications of Computer Science as we know them would not exist.

In the medical domain, data may be available in several forms, ranging from diagnostic reports written in natural language to electronic health records, multi-omics data, and so on. Among these, *medical images* constitute a large portion of the data that can be related to patients, or used for research purposes. The rise of Artificial-Intelligence (AI) based methods has widely augmented these datasets with computer-generated images (e.g. identifying lesions or regions of interest) or annotations. With respect to Medical Imaging, we mention a notable research effort towards querying and information retrieval (see the literature review in [19] and the citations therein). A major role in this area is currently played by *content-based information retrieval* (CBIR) [36, 31, 3, 20, 30].

The survey [25] mentions four key issues for research in the field, among which two are of interest to this paper: *the lack of effective representation of medical content by low-level mathematical features* and *the absence of appropriate tools for medical experts to experiment with a CBIR application*. In the present paper, we present a research line that aims at addressing these issues, among others, by the adoption of a coherent, user-oriented, expert-centric **declarative** computation paradigm. In doing so, we would like to emphasize an underlying problem when managing large, diverse medical imaging datasets: the **lack of a general-purpose query language for images**. Such a language should be able to identify regions of interest either by value, by imaging features (e.g. statistical texture analysis), and by spatial/topological characteristics (relative distance, contact, boolean operations, inter-reachability through other regions). Furthermore, it should make use of a diversity of data sources, ranging from patients datasets, to manually annotated ground truth, to the output of Machine Learning methods, and connect the information therein through expert-driven declarative queries and procedures.

Our proposed approach embraces a classic tenet of AI: that of *spatial logics*, that we shall discuss in the remainder of this work, and that constitutes the core of our initiative. More precisely, our work stems from the so-called *topological spatial logics* [1], where the object of reasoning are points, not regions (see [15] for an encoding of the *region calculus* of [33] into an extended topological spatial logic). Notably, we propose *model checking* to *automatically* identify sets of points (therefore, ultimately regions) satisfying user-specified properties. This

---

[4] See https://jquery.com

is similar to how other fully-automated methods, such as Machine Learning, are used nowadays, and **significantly diverges from *deduction***, which is the traditional approach of *Symbolic Reasoning* in AI.

Several publications related to spatial model checking, and in particular to healthcare-related applications, appeared recently. For instance, in [23], spiral electric waves – a precursor to atrial and ventricular fibrillation – are detected and specified using a spatial logic and model-checking tools. The formulas of the logic are learned from the spatial patterns under investigation and the onset of spiral waves is detected using bounded model checking. In [34] (see also the references in that paper), the authors describe mereotopological methods to programmatically correct image segmentation errors, exploiting a spatial logic called *discrete mereotopology* to integrate a number of qualitative spatial reasoning and constraint satisfaction methods into imaging procedures.

The group of authors of this paper have participated in joint publications on spatial logics and related model-checking approaches. The *Spatial Logic of Closure Spaces* (SLCS) has been defined in [13, 14], and used in several applications related to smart cities (see e.g. [12]). Recently, these methods were adapted to the efficient analysis of medical images based on Expert Knowledge [8, 9, 4]. The Free and Open Source spatial model checker VoxLogicA[5] has been developed to support an innovative methodology to analyse medical images. Such methodology obtained excellent experimental results. More precisely, in [9] the accurate contouring of brain tumour tissue obtained using VoxLogicA has been compared to the best performing algorithms (among which many based on deep learning) on a very relevant public benchmark data set for brain tumours (BraTS 2017 [35]). The obtained results are well in line with the state of the art, both in terms of accuracy and in terms of computational efficiency (on a related note, recent efforts have been devoted to running VoxLogicA on GPUs obtaining a substantial speed-up, see [11]). Furthermore, in [6, 7], VoxLogicA has been used for nevus segmentation, again obtaining results in line with the state of the art. It is also worth noting that the logic SLCS has been adopted, and extended, also in other contexts and by other groups of authors; for instance, for cyber-physical systems ([37]), or for run-time monitoring [32, 5].

The intended applications of VoxLogicA are not only novel autocontouring methods, but also the formalization of inter and intra-site workflows and collaboration patterns, and monitoring or *quality assurance* of autocontouring procedures, by encoding well-established protocols or guidelines. In future work, we aim at leveraging the spatial model checker VoxLogicA [9] as the distributed execution engine for an Open Platform for collaboration and data management in novel data-centric healthcare applications.

2   Spatial Logics for Medical Imaging

Our approach to the analysis of medical images is based on the fact that digital images can be seen as 2D or 3D regular grids, i.e. graphs where each node

---

[5] See https://github.com/vincenzoml/VoxLogicA

corresponds to a voxel and has a fixed number of adjacent nodes. The exact set of nodes adjacent to any given one depends on the particular *adjacency relation* between nodes one chooses. For instance, for 2D images, this set is composed of the pixel itself plus those other pixels with which it shares an edge, if the so called *orthogonal* adjacency relation is chosen, whereas it is composed of the pixel itself plus those other pixels with which it shares an edge or a vertex, if the *orthodiagonal* adjacency relation is considered. As these examples suggest, any adjacency relation must be a reflexive and symmetric binary relation over the nodes of the graph. Graphs, in turn, can be seen as a subclass of *closure spaces*, a generalisation of topological spaces; whenever the edge relation of the graph is an adjacency relation, we speak of *adjacency spaces*[6]. Thus, the theoretical foundations of our approach have their roots in topology and related notions [21]. It is convenient to associate each node of any such graph with some specific information, that can be represented as an *atomic predicate*, possibly expressing a property of some *attribute* of the node. For instance, in the case of black & white digital images, the relevant attribute of any voxel is the *intensity*, which has typically a value in the range $0 - 255$, whereas predicates of interest could be those expressing the fact that the voxel is in the border of the image, or that its intensity is lower than a certain threshold.

SLCS offers specific operators for reasoning about (points in) *closure models*, i.e. closure spaces enriched with atomic predicates. Besides general logical operators like conjunction ($\wedge$), disjunction ($\vee$) and negation ($\neg$), the most basic one for adjacency spaces is the *reachability operator* ρ; the formula $\rho\,\Phi_1\,[\Phi_2]$ is satisfied by a voxel x in an image $\mathcal{M}$ if there is a voxel t in $\mathcal{M}$ and a (possibly empty) sequence of *adjacent* voxels $x_1, \ldots, x_n$ in $\mathcal{M}$ such that x_1 is adjacent to x, x_n is adjacent to t that satisfies Φ_1 and all x_j satisfy Φ_2. For instance, $\rho\,\mathtt{red}[\mathtt{blue} \vee \mathtt{green}]$ means that we are interested in those voxels that can reach any red voxel through a sequence of points that must be blue or green. The *near* operator ($\mathcal{N}$) expresses the fact that any point satisfying $\mathcal{N}\,\Phi$ satisfies Φ or is adjacent to a point satisfying Φ; in fact, $\mathcal{N}\,\Phi$ is equivalent to $\rho\,\Phi\,[\bot]$. Similary, the formula $\Phi_1\,\mathcal{S}\,\Phi_2$, expressing that fact the the relevant point lays in an area the points of which satisfy Φ_1 and this area is *surrounded* by points satisfying Φ_2, is equivalent to $\Phi_1 \wedge \neg(\rho\,\neg(\Phi_1 \vee \Phi_2)[\neg\Phi_2])$.

The *Image Query Language* (ImgQL) comprises SLCS but is enriched with the distance operator $\mathcal{D}^I$ and the statistical similarity operator $\triangle$. Formula $\mathcal{D}^I\,\Phi$ is satisfied by any point x whose distance $d(x, \llbracket\Phi\rrbracket)$ from the set $\llbracket\Phi\rrbracket$ of points satisfying Φ falls in interval I. Intuitively, a point x satisfies the similarity operator $\triangle$ if a *region of interest* around x *correlates* with a *sample area*. More precisely, given a sample area $\llbracket\Phi\rrbracket$ specified by formula Φ and a region of interest defined as the *sphere* $S(x, r)$ of given radius r around x, the *cross-correlation*[7] between the histogram of an attribute a of the points in $S(x, r)$ and the histogram of

---

[6] For the purposes of this paper, the terms "voxel", "pixel", "node", and "point" can be considered synonims.

[7] In ImgQL, the normalised Pearson's correlation coefficient is used; 1 means perfect correlation, -1 means perfect anti-correlation, and 0 indicates no correlation.

an attribute b of the points in $\llbracket \Phi \rrbracket$ is compared with a given threshold c. The histogram of a given attribute of an image is a representation of the distribution of the values of the attribute; in practice, the range $[m, M]$ of interest of the attribute is split in a certain number k of adjacent intervals $B_1, \ldots, B_k$ of the same size, called *bins*. The histogram of the image is then the function H such that $H(j)$ gives the number of voxels with the value of that attribute falling within B_j. For example, $\triangle_{=1} \left[\begin{smallmatrix} 160 & 250 & 10 \\ 0.2 & \text{intensity} & \text{intensity} \end{smallmatrix} \right] \Phi$ identifies all the voxels whose neighbourhood of radius 0.2 units perfectly correlates with $\llbracket \Phi \rrbracket$; for the comparison, the chosen range of the intensity is $[160, 250]$ (for both the neighbourhoods and $\llbracket \Phi \rrbracket$) and has been split into 10 bins (of size 9 units). On the other hand, $\triangle_{<0.3} \left[\begin{smallmatrix} 160 & 250 & 10 \\ 0.2 & \text{intensity} & \text{intensity} \end{smallmatrix} \right] \Phi$ requires that the cross-correlation be less than 0.3.

3 The spatial model checker VoxLogicA

VoxLogicA is a framework for image analysis, that embeds the logic SLCS into ImgQL, a user-oriented expression language to manipulate images. The latter also includes additional selected imaging primitives. More precisely, the VoxLogicA type system distinguishes between *boolean-valued* images, that can be arguments or results of the application of SLCS operators, and *number-valued* images, resulting from imaging primitives. The underlying execution engine is a *global model checker*, that is, the set of all points satisfying a logic formula is computed at once. Functionality-wise, VoxLogicA specialises the former prototype spatio-temporal model checker topochecker[8] to the case of spatial analysis of *multi-dimensional images*. It interprets a specification using a set of multi-dimensional images[9] as models of the spatial logic, and produces as output a set of multi-dimensional images representing the valuation of user-specified expressions. For logical operators, such images are Boolean-valued, that is, *regions of interest* or *masks* in medical imaging terminology, which may be loaded as *overlays* in medical image viewers. Non-logical operators result in standard, number-valued images. Additionally, VoxLogicA offers file loading and saving primitives, and a set of additional operators, specifically aimed at image analysis, that is destined to grow along with future developments of the tool. The main execution modality of VoxLogicA is *batch execution*. A (currently experimental) *graphical user interface* is under development. A planned future development is *interactive execution*, in particular for semi-automated analysis, by letting a domain expert calibrate numeric parameters in real-time, while seeing the intermediate and final results. Implementation-wise, the tool achieves a two-orders-of-magnitude speedup with respect to topochecker. Such speedup has permitted the rapid development of a novel procedures for automatic segmentation that, besides being competitive with respect to the state-of-the-art in the field (see Section 4),

---

[8] See https://github.com/vincenzoml/topochecker

[9] Besides common bitmap formats, the model loader of VoxLogicA currently supports the NIfTI (Neuro-imaging Informatics Technology Initiative) format (https://nifti.nimh.nih.gov/, version 1 and 2). 3D MR-FLAIR images in this format very often have a slice size of 256 by 256 pixels, multiplied by 20 to 30 slices.

are also easily *replicable* and *explainable* to humans, and therefore amenable of improvement by the community of medical imaging practitioners.

VoxLogicA is free software and it is available in binary form for the operating systems Linux, OSX, and Windows.

4   Applications in Medical Imaging

4.1   Case study: Brain Tumour Segmentation

Glioblastoma multiforme (GBM) is the most common brain malignancy and is almost always lethal [43]. Survival after 2 years is achieved in only about 9% of patients. Medical images play a crucial role in the characterisation and in the treatment of the disease.

The first-line treatment of Glioblastoma is Surgery followed by radiotherapy. Crucial for radiotherapy is the accurate contouring of tissues and organs at risk, employing Magnetic Resonance (MR) and Computed Tomography (CT) images. Recent research efforts in the field have therefore been focused on the introduction of automatic or semi-automatic contouring procedures. More broadly speaking, such procedures can be used to identify particular kinds of tissues, such as parts of the brain (white matter, grey matter) or brain tumour related tissues. Such (semi-)automatic procedures would lead to an increase in accuracy and a considerable reduction in time and costs, compared to the current practice of manual contouring [38]. Automatic contouring of GBM is an open and challenging topic, since GBM is an intrinsically heterogeneous brain tumour, both in appearance, in shape, and in histology. The MICCAI Conference is organising a yearly challenge for brain tumour segmentation, since 2012, providing a common benchmark of brain lesion images, together with their ground truth segmentation approved by experienced neuro-radiologists, in the Brain Tumor Image Segmentation Benchmark (BraTS). One of our specifications in ImgQL has been validated in [9] using the 2017 BraTS dataset [28] containing multi-institutional pre-operative MRI scans of 210 patients affected by high grade gliomas. A priori, 17 cases have been excluded as they present multi-focal tumours or artifacts in the acquisition that the current procedure is not meant to deal with. The executable specification of the segmentation procedure consists of a concise, 30 lines long, text-file where the part concerning segmentation occupies only 10 lines, as shown in Specification 1, explained by the following steps:

1. Initial identification of the hyperintense regions (lines 2-5) in the MRI (of type FLAIR). These are areas with voxel intensity > 0.95 centile grown up to areas > 0.86 centile (growTum in line 6);
2. Search for voxels with a surrounding histogram similar (cross correlation > 0.6) to the area growTum (tumStatCC in line 8);
3. Identification of Gross Tumor Volume (GTV) by growing growTum up to the tumStatCC area. The GTV is then enlarged by 2.5 cm to simulate the Clinical Target Volume (CTV) in radiotherapy (line 9).

ImgQL Specification 1: Tumour segmentation method

```
 1 let pflair = percentiles(flair,brain,0)
 2 let hI = pflair >. 0.95
 3 let vI = pflair >. 0.86
 4 let hyperIntense = flt(5.0,hI)
 5 let veryIntense = flt(2.0,vI)
 6 let growTum = grow(hyperIntense,veryIntense)
 7 let tumSim = similarTo(5,growTum,flair)
 8 let tumStatCC = flt(2.0,(tumSim >. 0.6))
 9 let gtv= grow(growTum,tumStatCC)
10 let ctv = distlt(25,gtv) & brain
```

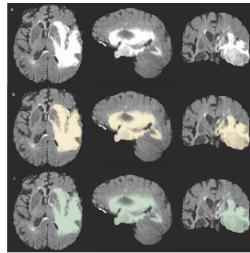Figure 1 shows an example of the segmentation, together with the ground truth.

Fig. 1: GTV for patient TCIA_471 from the BraTS 2017 dataset. Top row: original image. Middle row: ground truth. Bottom row: `VoxLogicA` segmentation.

The segmentation results for the Brats 2017 benchmark are reproduced from [9] in Table 1 where they are compared to the ground truth images based on the common similarity coefficients Dice, Sensitivity and Specificity for the GTV and CTV areas. The 3D images consist of $240 \times 240 \times 155$ voxels (ca. 9 M voxels) and the evaluation for each patient takes about 10 seconds on a desktop computer with an Intel Core I7 7700 processor and 16GB of RAM.

	Sensitivity (193 cases)	Specificity (193 cases)	Dice (193 cases)	Sensitivity (210 cases)	Specificity (210 cases)	Dice (210 cases)
GTV	0.89(0.10)	1.0(0.00)	0.85(0.10)	0.86(0.16)	1.0(0.0)	0.81(0.18)
CTV	0.95(0.07)	0.99(0.01)	0.90(0.09)	0.93(0.14)	0.99(0.2)	0.87(0.15)

Table 1: `VoxLogicA` evaluation on the BraTS 2017 data set. Adapted from [9].

4.2   Case study: Nevus Segmentation

Melanoma is the most serious form of skin cancer, the incidence of which has been increasing for many decades [26, 22]. While the disease may be lethal, its correct and early detection, and its consequent treatment, results in no change in life expectancy [29]. Disease recognition is performed through *dermoscopy*, a specialized technique of high-resolution imaging of the skin, allowing specialists to visualize deeper underlying structures of the skin lesions. This technique has been proved to have diagnostic accuracy of up to 84% when carried out by specially trained clinicians [2]. However, one of the main issues is that in many countries, there is only a limited number of such specialists available. Therefore, there is a clear need for automated methods that can help to recognise the disease reliably and at an early stage so that more lives could be saved [17].

The International Skin Imaging Collaboration (ISIC) is a collaboration between academia and industry to find automatic techniques to detect melanoma from dermoscopy images. Since 2016 ISIC organizes challenges titled "Skin Lesion Analysis toward Melanoma Detection" [17]. The first task involved in the challenges, and more in general in the diagnosis of melanoma, is the skin lesion segmentation. In [6] we investigate the feasibility of the application of a procedure implemented in ImgQL for the segmentation of images of nevi from two datasets released by ISIC for the 2016 challenge: a *training* set and a *test* set of 900 and 379 images, respectively. Both datasets contain annotated dermoscopic images and the corresponding ground truth (i.e. a segmentation performed manually by experts) for each image. One of the challenges with such datasets, and, more in general, with dermoscopic images of skin lesions, is their great inhomogeneity. Nevi may show nonuniform colour ranges, their colour may have more or less contrast with the colour of skin, may have different sizes, may have more or less smooth borders, and may be composed of different parts; moreover, the skin may be more or less regular, with the presence of hairs or sebaceous follicles. Furthermore, images may also show heterogeneity due to the dermoscope used: they may be of different size, showing black corners, rings, shadows, or ultrasound gel drops, showing more or less contrast and intensity. Finally, images may also show extraneous elements such as patches or ink marks. The images in Figure 2 show a few examples of this inhomogeneity.

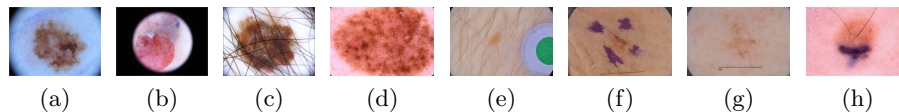(a)   (b)   (c)   (d)   (e)   (f)   (g)   (h)

Fig. 2: Example images from the ISIC datasets illustrating the inhomogeneity of nevi.

Due to such great variability in the dermoscopic images datasets, our procedure starts from two basic assumptions: (at least part of) the nevus is in the

middle of the image; pixels belonging to the skin are close to the border of the image. Our aim is to distinguish skin tissue from nevus tissue with the help of the texture analysis and other spatial operators. The executable procedure consists of less than 30 lines of code and consists essentially of five parts:

1. *Identify a sample of the skin.*
   First, the black corners that can be observed in many images (see for instance Fig. 2b) are identified as all the dark pixels having an intensity below 40 that can reach the border of the image exclusively 'passing by' further dark pixels. Then all pixels that are at most at relative distance 200 from the black borders of the image are taken. Due to the differences in the size of the images all the distances are scaled taking for reference an image of 1022 pixels of width and 767 pixels of height.
   Patches present in some of the images (see for instance Fig. 2e) are excluded from the sample of the skin. Patches are identified using the information about their colour (or their brightness in case of white patches), their size (they cover less than 40% of the image), their position (they are at the border of the image), and their compactness.

2. *Identify pixels belonging to the skin.*
   The similarity score of each pixel in the image with respect to the sample of the skin identified in the previous phase is computed using the similarity operator. Pixels with a cross-correlation higher than 0.05 are considered skin.

3. *Preliminary nevus segmentation.*
   To identify a preliminary segmentation of the nevus we use the information about the area around the nevus, where some of the pixels belong to the nevus and some to the healthy skin. From the histogram of the intensity of the image resulting from the similarity score computed in the second phase, we can identify the value which works like a watershed between pixels that are part of the nevus and pixels that are part of the skin.

4. *Final nevus segmentation.*
   The segmentation obtained so far is extended with points that are sufficiently similar to the nevus.

5. *Comparing the segmentation with the ground truth.*
   The segmentation obtained (nevSegm) is compared with the ground truth provided by the ISIC 2016 challenge for both the *training* and *test* sets. For this comparison common similarity indexes were used: Dice, Jaccard, Accuracy, Sensitivity (SE), and Specificity (SP). Table 2 shows the *mean* values for these indexes.

The images in Figure 3 show the resulting images for each step in the segmentation procedure of image ISIC_0008294

5   Querying medical image datasets

As of today, the toolkit of a Medical Imaging practitioner mostly consists of traditional programming languages, such as the ubiquitous *python*[10]. Program-

---

[10] See https://www.python.org/

	Accuracy	Dice	Jaccard	SE	SP
nevSegm: Mean *Training* set	0.902	0.818	0.726	0.810	0.965
nevSegm: Mean *Test* set	0.899	0.809	0.717	0.802	0.960

Table 2: Average similarity scores of nevSegm for images of the ISIC 2016 *training* set (all 900 images) and *test* set (all 379 images).

(a)   (b)   (c)   (d)

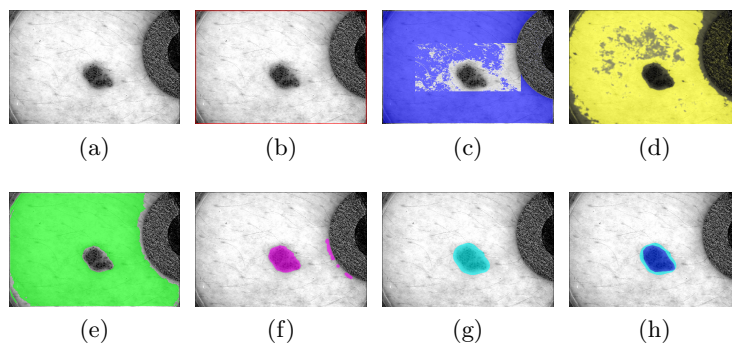(e)   (f)   (g)   (h)

Fig. 3: Segmentation of image ISIC_0008294. Figure (a) shows the nevus intensities (greyscale); (b) shows the border of the image (red); (c) shows the sample of the skin (blue); (d) shows the similarity score of each pixels w.r.t. the sample of the skin (variation of the score is shown as the varying intensity of yellow); (e) shows pixels with a similarity higher than 0.05 w.r.t. the sample of the skin (green); (f) shows the pre segmentation of the nevus (magenta); (g) shows the final nevus segmentation (cyan); (h) shows the comparison between the automatic segmentation (cyan) and the ground truth (blue).

ming languages are used in conjunction with libraries of imaging primitives[11], and with Machine Learning (ML) libraries, such as *Keras*[12].

Although such a setup is quite well-established from a programmer's perspective, it is not meant to be used by non-programmers. Domain experts such as Medical Doctors (for instance: Radiotherapists), MRI technologists, Medical Physicists, Healthcare researchers, and even end-users lack a general ability to interrogate medical imaging databases, for instance, in order to search for specific features in images, compose results from different methods to explore the design space of novel analysis techniques, or evaluate the impact of changes in an imaging workflow.

---

[11] See, for instance, the *Insight Toolkit (ITK)*, https://itk.org/, also used under-the-hood in `VoxLogicA`, and the *Open Source Computer Vision Library (OpenCV)*, https://opencv.org/.

[12] See https://keras.io/.

It is worth noting that, even if ML is nowadays a widespread methodology in (medical) image analysis, there are a number of interesting problems that, as of today, are not meant not be addressed using ML alone.

Example 1. As a concrete example, consider **clinical guidelines** (see e.g. [18]). These usually consist of a series of steps and checks that the Radiotherapist should execute in order to get to an acceptable contouring of a tumour. A hospital may want to perform monitoring and routine checks, and to prepare for changing such guidelines when new ones are published. Before embracing a new set of guidelines, in order to evaluate their impact, the Radiotherapy department might want to investigate how much their standard practice obeys to the new guideline, and how much in practice the old and new guidelines differ on their own patients. A language is needed that can express such guidelines in a concise, human-friendly, expert-oriented way, and that can execute them automatically, highlighting their interpretation on patient images. ML alone cannot currently be used for such tasks, as the encoding of guidelines is, more broadly speaking, a matter of *expert knowledge*. In [42], a model checker is used for the purpose.

Example 2. A declarative, unambiguous, executable encoding of specific guidelines would also be extremely important for **quality assurance** of ML-based methods (e.g., autocontouring for Radiotherapy), enabling those who are responsible for the treatment to monitor and assess the operation of such algorithms in accordance with selected protocols.

Example 3. A research-oriented example is that of **identifying relevant cases to test a research hypothesis**. Imagine for instance, that a researcher needs to check a novel hypothesis relating a spatial feature (say: "the proximity of a brain tumour to the cerebellum"), with an aspect of the treatment (e.g., the survival rate of patients, or the outcome of radiotherapy). For this, it is necessary to filter an existing dataset to find all the cases that satisfy the hypothesis. This is a kind of spatial query on images that cannot be carried on using ML, for the simple reason that there is no training data for detecting such a specific feature as "the tumour is very close to the cerebellum". Traditional programming can be used to identify such cases. However, writing a full program for each such query is not only time consuming, but also practically impossible for researchers who are not expert programmers. Also note that it is not guaranteed that such a program would be efficient enough to analyse the full dataset under time constraints. A specifically designed, declarative, optimizing query engine such as `VoxLogicA` is better suited for the task. Techniques such as *memoization*[39], automatic parallel execution (see [9]), on-disk caching, and *graphical processing units* (see [11]), are used to speed up computation.

Example 4. Healthcare authorities might want to **evaluate the potential impact of new therapies**, by identifying the number of patients that may get benefits. Imaging features could be used to identify such cases; consider e.g. the case of tumours invading specific organs at risk. For instance, consider the query: "find all the cases in which a brain tumour invaded the patient's eye".

Once again, there is no training data specific to this query, even if it is likely that Machine Learning will provide in the near future very good methods to identify the simpler concepts of "brain tumour", and "patient's eye'. In this case, a query language would be an excellent solution to coordinate different machine learning algorithms in order to answer more complex, expert-driven questions.

Example 5. An easy-to-use query system for healthcare datasets definitely creates ethical concerns about its usage. But at the same time, the human readability of the query language makes it *transparent* with respect to such concerns and would ease **ethical scrutiny** when deemed necessary, giving to an ethics committee the ability to know exactly *which questions have been asked to the system, and what do they mean.* Compare and contrast this to a *black-box* method, such as ML; or to the complexity of a traditional computer program interfaced to a database of clinical data.

Example 6. **Remote collaboration in telemedicine** could benefit of an automated system to identify and share regions of interest among several treatment centres. In this case, the flexibility of a query language would be fundamental to integrate different queries into a comprehensive *personalized knowledge base* for each patient.

6   Outlook: `VoxLogicA` as a distributed query engine

In our opinion, the tool `VoxLogicA` as is, would already constitute a good starting point for querying medical imaging datasets (and, possibly, their augmentation with information derived from ML algorithms). However, we also take into account a number of improvements, briefly introduced below, which could direct future research more closely to the field of Information Retrieval, in order to be used in the way that we anticipated throughout Examples 1 – 6 of the previous section, namely:

Integration with clinical databases. Information systems such as *RIS-PACS* [24] are of vital importance for the functioning of modern healthcare. Therein, all the information available about patients that undergo treatments is archived. In a clinical setting, therefore, a query language for medical imaging would be way more useful if enabled to query such information systems directly.

Indexing. In SQL databases, data is indexed in order to speed up queries. In `VoxLogicA`, the most expensive queries are related to the computation of distance maps, to connected components and to statistical texture analysis. Therefore, the identification of data structures to be computed in advance from medical images (e.g., when such images are added to a dataset) in order to speed up such operations, would be a relevant research line. In this respect, a first step is to observe that *minimization* may be used to reduce models up-to *logical equivalence.* In [16], and more recent work (to appear), a minimization algorithm is proposed that minimizes images up-to proximity ("near") and reachability queries.

Distributed execution. Distributed execution of VoxLogicA queries could be very relevant for broadly known, widespread datasets, such as those that are commonly used for research and benchmarking purposes (see e.g. the BraTS dataset [35]). Also for multi-centric studies, in order to respect privacy and intellectual property concerns, it would be worthwile to implement a fully distributed variant of VoxLogicA that can be used to interrogate remote medical systems and draw statistical conclusions, without having to share the whole dataset across all the participants in a study.

Computational Efficiency. In order to improve efficiency of queries, the strategy that is currently under investigation is that of GPU computing. The current on-GPU implementation of `VoxLogicA` [11] exploits Graphical Processing Units to improve the performance of image analysis. As software portability is a major issue, VoxLogicA GPU is currently implemented using OpenCL[13], an open standard by *Kronos*, known to be executable on GPUs from any vendor. The prototype is currently being improved. A major bottle-neck is the computation of *Connected Components* of a binary image; this is used to resolve inter-reachability queries on image regions, and it requires iterative calls to the GPU and comparisons between images, which is particularly computationally intensive. However, the speed-up achieved by the propotype is quite substantial, and may become as high as two orders of magnitude, depending on the size and type of the evaluated formula.

Human-computer interaction. The experience of our group in using `VoxLogicA` for medical image analysis against a large datasets, is that a major hurdle is constituted by HCI issues related to the visualization of intermediate results, the interactive construction of queries, and to comparing and exchanging several different versions of the same analysis between a group of interested users. Furthermore, the target group of users for our project is that of healthcare practitioners, where user interfaces for image visualization employ some highly standardised concepts (such as viewing images slice by slice, using *axial*, *coronal* and *sagittal* projections, or using overlays on images to visualize regions of interest). Such aspects need to be carefully evaluated in order to design a query system for medical images which is effective (see e.g. the usability study in [27], and the cognitive load issues investigated in [10]).

References

1. Aiello, M., Pratt-Hartmann, I., Benthem, van, J.: Handbook of Spatial Logics. Springer (2007)
2. Ali, A.R.A., Deserno, T.M.: A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data. In: Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment. vol. 8318, p. 83181I. International Society for Optics and Photonics (2012)

---

[13] See https://www.khronos.org/opencl/.

3. Aswani, N., Beckers, T., Birngruber, E., Boyer, C., Burner, A., Bystron, J., Choukri, K., Cruchet, S., Cunningham, H., Dedek, J., et al.: Khresmoi: multimodal multilingual medical information search. In: Proceedings of the 24th international conference of the European federation for medical informatics. p. 12 (2012)

4. Banci Buonamici, F., Belmonte, G., Ciancia, V., Latella, D., Massink, M.: Spatial logics and model checking for medical imaging. Int. J. Softw. Tools Technol. Transf. 22(2), 195–217 (2020), https://doi.org/10.1007/s10009-019-00511-9

5. Bartocci, E., Bortolussi, L., Loreti, M., Nenzi, L., Silvetti, S.: Moonlight: A lightweight tool for monitoring spatio-temporal properties. In: Deshmukh, J., Nickovic, D. (eds.) Runtime Verification - 20th International Conference, RV 2020, Los Angeles, CA, USA, October 6-9, 2020, Proceedings. Lecture Notes in Computer Science, vol. 12399, pp. 417–428. Springer (2020), https://doi.org/10.1007/978-3-030-60508-7_23

6. Belmonte, G., Broccia, G., Ciancia, V., Latella, D., Massink, M.: Using spatial logic and model checking for nevus segmentation. CoRR abs/2012.13289 (2020), https://arxiv.org/abs/2012.13289

7. Belmonte, G., Broccia, G., Vincenzo, C., Latella, D., Massink, M.: Feasibility of spatial model checking for nevus segmentation. In: FORMALISE: International Conference on Formal Methods in Software Engineering. p. To appear (2021)

8. Belmonte, G., Ciancia, V., Latella, D., Massink, M.: From collective adaptive systems to human centric computation and back: Spatial model checking for medical imaging. In: ter Beek, M.H., Loreti, M. (eds.) Proceedings of the Workshop on FORmal methods for the quantitative Evaluation of Collective Adaptive SysTems, FORECAST@STAF 2016, Vienna, Austria, 8 July 2016. EPTCS, vol. 217, pp. 81–92 (2016), https://doi.org/10.4204/EPTCS.217.10

9. Belmonte, G., Ciancia, V., Latella, D., Massink, M.: Voxlogica: A spatial model checker for declarative image analysis. In: Vojnar, T., Zhang, L. (eds.) Tools and Algorithms for the Construction and Analysis of Systems - 25th International Conference, TACAS 2019, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2019, Prague, Czech Republic, April 6-11, 2019, Proceedings, Part I. Lecture Notes in Computer Science, vol. 11427, pp. 281–298. Springer (2019), https://doi.org/10.1007/978-3-030-17462-0_16

10. Broccia, G., Milazzo, P., Ölveczky, P.C.: Formal modeling and analysis of safety-critical human multitasking. Innov. Syst. Softw. Eng. 15(3-4), 169–190 (2019), https://doi.org/10.1007/s11334-019-00333-7

11. Bussi, L., Ciancia, V., Gadducci, F.: A spatial model checker in GPU (extended version). CoRR abs/2010.07284 (2020), https://arxiv.org/abs/2010.07284

12. Ciancia, V., Gilmore, S., Grilletti, G., Latella, D., Loreti, M., Massink, M.: Spatio-temporal model checking of vehicular movement in public transport systems. International Journal on Software Tools for Technology Transfer (Jan 2018), https://doi.org/10.1007/s10009-018-0483-8

13. Ciancia, V., Latella, D., Loreti, M., Massink, M.: Specifying and verifying properties of space. In: Theoretical Computer Science - 8th IFIP TC 1/WG 2.2 International Conference, TCS 2014, Rome, Italy, September 1-3, 2014. Proceedings. Lecture Notes in Computer Science, vol. 8705, pp. 222–235. Springer (2014)

14. Ciancia, V., Latella, D., Loreti, M., Massink, M.: Model Checking Spatial Logics for Closure Spaces. Logical Methods in Computer Science Volume 12, Issue 4 (Oct 2016), http://lmcs.episciences.org/2067

15. Ciancia, V., Latella, D., Massink, M.: Embedding RCC8D in the Collective Spatial Logic CSLCS. In: Boreale, M., Corradini, F., Loreti, M., Rosario, P. (eds.)

To be defined, pp. 251–266. Lecture Notes in Computer Science, Springer Berlin Heidelberg (2019), accepted for publication

16. Ciancia, V., Latella, D., Massink, M., de Vink, E.: Towards spatial bisimilarity for closure models: Logical and coalgebraic characterisations (2020)

17. Codella, N.C., Nguyen, Q.B., Pankanti, S., Gutman, D.A., Helba, B., Halpern, A.C., Smith, J.R.: Deep learning ensembles for melanoma recognition in dermoscopy images. IBM Journal of Research and Development 61(4/5), 5–1 (2017), http://www.research.ibm.com/journal/, special Issue on Deep Learning

18. Combs, S.E., Baumert, B.G., Bendszus, M., Bozzao, A., Brada, M., Fariselli, L., Fiorentino, A., Ganswindt, U., Grosu, A.L., Lagerwaard, F.L., Niyazi, M., Nyholm, T., Paddick, I., Weber, D.C., Belka, C., Minniti, G.: ESTRO ACROP guideline for target volume delineation of skull base tumors. Radiotherapy and Oncology 156, 80–94 (Mar 2021), https://doi.org/10.1016/j.radonc.2020.11.014

19. Demner-Fushman, D., Müller, H., Kalpathy-Cramer, J., Antani, S.: A decade of community-wide efforts in advancing medical images understanding and retrieval. Computerized Medical Imaging and Graphics (2014), http://publications.hevs.ch/index.php/attachments/single/674

20. Deserno, T.M., Güld, M.O., Thies, C., Plodowski, B., Keysers, D., Ott, B., Schubert, H.: Irma - content-based image retrieval in medical applications. Studies in health technology and informatics 107 Pt 2, 842–6 (2004)

21. Galton, A.: A generalized topological view of motion in discrete space. Theor. Comput. Sci. 305(1-3), 111–134 (2003), https://doi.org/10.1016/S0304-3975(02)00701-6

22. Godar, D.E.: Worldwide increasing incidences of cutaneous malignant melanoma. Journal of skin cancer 2011 (2011)

23. Grosu, R., Smolka, S., Corradini, F., Wasilewska, A., Entcheva, E., Bartocci, E.: Learning and detecting emergent behavior in networks of cardiac myocytes. Commun. ACM 52(3), 97–105 (2009)

24. Haux, R.: Strategic information management in hospitals : an introduction to hospital information systems. Springer, New York (2003)

25. Long, L.R., Antani, S., Deserno, T.M., Thoma, G.R.: Content-based image retrieval in medicine: Retrospective assessment, state of the art, and future directions. International journal of healthcare information systems and informatics : official publication of the Information Resources Management Association 4(1), 1–16 (Jan 2009), https://doi.org/10.4018/jhisi.2009010101, 20523757[pmid]

26. Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S.Y., et al.: Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. The lancet 380(9859), 2095–2128 (2012)

27. Markonis, D., Holzer, M., Baroz, F., Castaneda, R.L.R.D., Boyer, C., Langs, G., Müller, H.: User-oriented evaluation of a medical image retrieval system for radiologists. Int. J. Medical Informatics 84(10), 774–783 (2015), https://doi.org/10.1016/j.ijmedinf.2015.04.003

28. Menze, B.H.e.a.: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging 34(10), 1993–2024 (2015)

29. Mocellin, S., Nitti, D.: Cutaneous melanoma in situ: translational evidence from a large population-based study. The oncologist 16(6), 896 (2011)

30. Müller, H., Clough, P., Hersh, W., Deselaers, T., Lehmann, T., Geissbuhler, A.: Evaluation axes for medical image retrieval systems: The imageclef experience. In:

Proceedings of the 13th Annual ACM International Conference on Multimedia. p. 1014–1022. MULTIMEDIA '05, Association for Computing Machinery, New York, NY, USA (2005), https://doi.org/10.1145/1101149.1101358

31. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. International Journal of Medical Informatics 73(1), 1–23 (2004), https://www.sciencedirect.com/science/article/pii/S1386505603002119

32. Nenzi, L., Bortolussi, L., Ciancia, V., Loreti, M., Massink, M.: Qualitative and Quantitative Monitoring of Spatio-Temporal Properties with SSTL. Logical Methods in Computer Science 14(4), 1–38 (2018), DOI 10.23638/LMCS-14(4:2)2018. Published on line: 23 Oct. 2018. ISSN: 1860-5974

33. Randell, D.A., Cui, Z., Cohn, A.G.: A spatial logic based on regions and connection. In: Nebel, B., Rich, C., Swartout, W.R. (eds.) Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning (KR'92). Cambridge, MA, USA, October 25-29, 1992. pp. 165–176. Morgan Kaufmann (1992)

34. Randell, D.A., Galton, A., Fouad, S., Mehanna, H., Landini, G.: Mereotopological correction of segmentation errors in histological imaging. Journal of Imaging 3(4) (2017), https://www.mdpi.com/2313-433X/3/4/63

35. Spyridon (Spyros) Bakas et al. (Ed.): 2017 international MICCAI BraTS Challenge: pre-conference proceedings (Sept 2017), https://www.cbica.upenn.edu/sbia/Spyridon.Bakas/MICCAI_BraTS/MICCAI_BraTS_2017_proceedings_shortPapers.pdf

36. Tang, L.H., Hanka, R., Ip, H.H.: A review of intelligent content-based indexing and browsing of medical images. Health Informatics Journal 5(1), 40–49 (1999)

37. Tsigkanos, C., Kehrer, T., Ghezzi, C.: Modeling and verification of evolving cyber-physical spaces. In: Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering. pp. 38–48. ESEC/FSE 2017, ACM, New York, NY, USA (2017), http://doi.acm.org/10.1145/3106237.3106299

38. Vorwerk, H., Zink, K., Schiller, R., Budach, V., Böhmer, D., Kampfer, S., Popp, W., Sack, H., Engenhart-Cabillic, R.: Protection of quality and innovation in radiation oncology: The prospective multicenter trial the german society of radiation oncology (DEGRO-QUIRO study). Strahlentherapie und Onkologie 190(5), 433–443 (Mar 2014), https://doi.org/10.1007/s00066-014-0634-0

39. Wikipedia contributors: Memoization — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Memoization&oldid=1002117560 (2021), [Online; accessed 29-March-2021]

40. Wikipedia contributors: Sql — Wikipedia, the free encyclopedia (2021), https://en.wikipedia.org/w/index.php?title=SQL&oldid=1010996297, [Online; accessed 22-March-2021]

41. Wikipedia contributors: Xml — Wikipedia, the free encyclopedia (2021), https://en.wikipedia.org/w/index.php?title=XML&oldid=1012598275, [Online; accessed 22-March-2021]

42. Yousef Sanati, M., MacCaull, W., Maibaum, T.S.E.: Analyzing clinical practice guidelines using a decidable metric interval-based temporal logic. In: Jones, C., Pihlajasaari, P., Sun, J. (eds.) FM 2014: Formal Methods. pp. 611–626. Springer International Publishing, Cham (2014)

43. Zong, H., Verhaak, R.G., Canoll, P.: The cellular origin for malignant glioma and prospects for clinical advancements. Expert Review of Molecular Diagnostics 12(4), 383–394 (May 2012), https://doi.org/10.1586/erm.12.30