



NILDE

Network Inter-Library Document Exchange

Il presente documento viene fornito attraverso il servizio NILDE dalla Biblioteca fornitrice, nel rispetto della vigente normativa sul Diritto d'Autore (Legge n.633 del 22/4/1941 e successive modifiche e integrazioni) e delle clausole contrattuali in essere con il titolare dei diritti di proprietà intellettuale.

La Biblioteca fornitrice garantisce di aver effettuato copia del presente documento assolvendo direttamente ogni e qualsiasi onere correlato alla realizzazione di detta copia.

La Biblioteca richiedente garantisce che il documento richiesto è destinato ad un suo utente, che ne farà uso esclusivamente personale per scopi di studio o di ricerca, ed è tenuta ad informare adeguatamente i propri utenti circa i limiti di utilizzazione dei documenti forniti mediante il servizio NILDE.

La Biblioteca richiedente è tenuta al rispetto della vigente normativa sul Diritto d'Autore e in particolare, ma non solo, a consegnare al richiedente un'unica copia cartacea del presente documento, distruggendo ogni eventuale copia digitale ricevuta.

Biblioteca richiedente: CNR Biblioteca Area della Ricerca di Pisa
Data richiesta: 20/12/2021 14:50:18
Biblioteca fornitrice: Università di Milano Bicocca Biblioteca di Ateneo - Sede di Scienze
Data evasione: 21/12/2021 14:34:58

Titolo rivista/libro: Communications in computer and information science
Titolo articolo/sezione: Querying medical imaging datasets using spatial logics (Position paper)
Autore/i: Belmonte G., Broccia G.; Bussi L.; Ciancia V.; Latella D.; Massink M.
ISSN: 1865-0937
DOI: 10.1007/978-3-030-87657-9_22
Anno: 2021
Volume: 1481
Fascicolo:
Editore: Springer
Pag. iniziale: 285
Pag. finale: 301



Querying Medical Imaging Datasets Using Spatial Logics (Position Paper)

Gina Belmonte¹, Giovanna Broccia², Laura Bussi^{2,3}, Vincenzo Ciancia^{2(✉)},
Diego Latella², and Mieke Massink²

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
`vincenzo.ciancia@isti.cnr.it`

³ Dipartimento di Informatica, Università di Pisa, 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, in this paper we sketch a research plan devoted to the definition and implementation 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

The names of the authors of this paper are listed in alphabetical order. All co-authors have contributed equally to the work described herein. This work has been partially supported by the Italian MIUR-PRIN 2017 project IT MaTTerS: Methods and Tools for Trustworthy Smart Systems” and by the POR FESR 2014-2020 project STINGRAY (SmarT station INtelliGent RAILwaY).

© Springer Nature Switzerland AG 2021

L. Bellatreche et al. (Eds.): MEDI 2021 Workshops, CCIS 1481, pp. 285–301, 2021.

https://doi.org/10.1007/978-3-030-87657-9_22

us just name a few. The *Structured Query Language SQL* [12] revolutionised data representation and access, and is nowadays one of the pillars of modern Information Technology. The *eXtensible Markup Language XML*¹ and its query/transformation languages (see e.g. *XQuery* [1] and siblings), provide solid grounds to any modern data-centric or document-centric infrastructure. The javascript library *jQuery*² 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 [21] and the citations therein). A major role in this area is currently played by *content-based information retrieval* (CBIR) [4, 22, 33, 34, 39].

The survey [27] 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* [2], where the object of reasoning are points, not regions

¹ See <https://www.w3.org/XML/>.

² See <https://jquery.com>.

(see [16] for an encoding of the *region calculus* of [36] into an extended topological spatial logic). Notably, we propose *model checking to automatically* identify sets of points (therefore, ultimately regions) satisfying user-specified properties. Model-checking (in its “global” variant that we employ) is a fully automated technique that, taking as input a logical specification and a model of a system (in our case such “model” is just a digital image), returns the set of states (in our case, points) that satisfy the specification. This 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 Artificial Intelligence.

Several publications related to spatial model checking, and in particular to healthcare-related applications, appeared recently. For instance, in [25], 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 [37] (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 [14, 15], and used in several applications related to smart cities (see e.g. [13]). Recently, these methods were adapted to the efficient analysis of medical images based on Expert Knowledge [5, 8, 9]. The Free and Open Source spatial model checker **VoxLogicA**³ 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 [38]). 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 [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 ([40]), or for run-time monitoring [6, 35].

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

³ See <https://github.com/vincenzoml/VoxLogicA>.

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 a digital image can be seen as a 2D or 3D regular grid, i.e. a graph where each node corresponds to a pixel/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*⁴. Thus, the theoretical foundations of our approach have their roots in topology and related notions [23]. 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 express the fact that the voxel is in the border of the image, or that its intensity is lower than a certain threshold.

For the full syntax and semantics of SLCS, we refer the reader to [9, 15]. Here we provide an intuitive description of the logical language. 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, \dots, 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, $\rho\text{red}[\text{blue} \vee \text{green}]$ 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 $\rho\Phi[\perp]$. Similarly, 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 $\Phi_1 \wedge \neg(\rho\neg(\Phi_1 \vee \Phi_2)[\neg\Phi_2])$.

⁴ For the purposes of this paper, the terms “voxel”, “pixel”, “node”, and “point” can be considered synonyms.

The *Image Query Language* (ImgQL) comprises SLCS but is enriched with several imaging primitives, among which, more prominently, the Euclidean distance operator \mathcal{D}^I and the statistical similarity operator Δ . Formula $\mathcal{D}^I \Phi$ is satisfied by any point x whose Euclidean 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 Δ 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*⁵ between the histogram of an attribute a of the points in $S(x, r)$ and the histogram of an attribute b of the points in $\llbracket \Phi \rrbracket$ is compared with a given threshold c (see [5, 9] for further details).

3 The Spatial Model Checker VoxLogicA

VoxLogicA is a model checker for ImgQL. The VoxLogicA type system distinguishes between *boolean-valued* images, that usually are 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. The computational complexity of the procedure is linear in the number of sub-formulas times the number of points of the image [15], whereas efficiency-wise, in most cases the computation runs in a matter of seconds, on a high-end desktop computer [9]. Functionality-wise, VoxLogicA specialises the former prototype spatio-temporal model checker **topochecker**⁶ to the case of spatial analysis of *multi-dimensional images*⁷ 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

⁵ In ImgQL, the normalised Pearson's correlation coefficient is used; 1 means perfect correlation, -1 means perfect anti-correlation, and 0 indicates no correlation.

⁶ See <https://github.com/vincenzoml/topochecker>.

⁷ 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.

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 Sect. 4), 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

In this section, we exemplify the expressive power of the logical language used by `VoxLogicA` by reporting on two case studies. In both cases, only one image per case of the dataset is required to carry on the analysis; however, the tool is also capable of loading multiple images (say, an MRI image and a co-registered CAT image), resulting in multi-modal imaging capabilities. By these examples, we aim at demonstrating that the primitives of spatial logics are very close to the domain of discourse in medical imaging, by the simplicity and conciseness of the presented analysis procedures, that have reached state-of-the-art accuracy.

4.1 Case Study: Brain Tumour Segmentation

Glioblastoma multiforme (GBM) is the most common brain malignancy and is almost always lethal [44]. 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 [41]. 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 [31] 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

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

```

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) (Fig. 1).

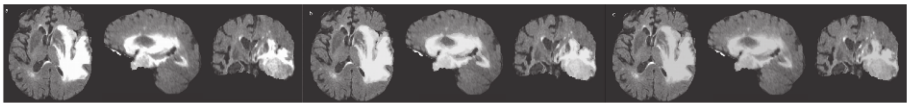


Fig. 1. GTV for patient TCIA_471 from the BraTS 2017 dataset. Leftmost: original image, where the tumour appears hyper-intense. Middle: ground truth in yellow (manually annotated by human experts). Rightmost: **VoxLogicA** segmentation result of Specification 1 in green (fully automated). (Color figure online)

The segmentation results for the Brats 2017 benchmark are reproduced from [9] in Table 1. The table shows the similarity scores for three commonly used coefficients, namely Dice, Sensitivity and Specificity, for the GTV and CTV areas. These scores indicate the similarity between the area segmented with **VoxLogicA** Specification 1 compared to the manually annotated images by human experts (i.e. ground truth images). Considering intra- and inter-experts variability of $20 \pm 15\%$ and $28 \pm 12\%$, respectively, for manual segmentations [30], Dice similarity scores above 0.8 are considered as very good, and above 0.9 as excellent. The 3D images consist of $240 \times 240 \times 155$ voxels (ca.

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

	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)

9 M voxels) and the evaluation for each patient takes about 5s on a desktop computer with an Intel Core I7 7700 processor and 16 GB of RAM.

Once a segmentation of the tumour has been obtained, it can also be used to select images in a dataset, based on some features of interest of the tumour area. For example, one may be interested in images in which the tumour has a volume (in terms of number of voxels it covers) that is larger than a certain threshold (VoxLogicA includes an operator for that). See Sect. 5 for more details.

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 [24,28]. While the disease may be lethal, its correct and early detection, and its consequent treatment, results in no change in life expectancy [32]. 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 [3]. 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 [19].

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” [19]. The first task involved in the challenges, and more in general in the diagnosis of melanoma, is the skin lesion segmentation. In [7] we investigated 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

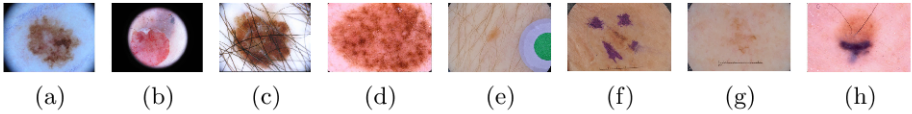


Fig. 2. Examples from the ISIC datasets illustrating the inhomogeneity of nevi.

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 Fig. 2 show a few examples of this inhomogeneity.

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* by removing the black borders and then taking a small region around the centre of the resulting image.
2. *Identify pixels belonging to the skin* using texture similarity with respect to the previously identified sample.
3. *Preliminary nevus segmentation* via a threshold on the image histogram.
4. *Final nevus segmentation* expanding the result of the previous step using texture similarity.

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 Fig. 3 show the resulting images for each step in the segmentation procedure of image ISIC_0008294.

Also in this case, after relevant features of nevi have been identified using voxel-based spatial logics, VoxLogicA can be used to query a dataset in order to find images that have specific features (e.g. irregular contour, or large size). See also Sect. 5.

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

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

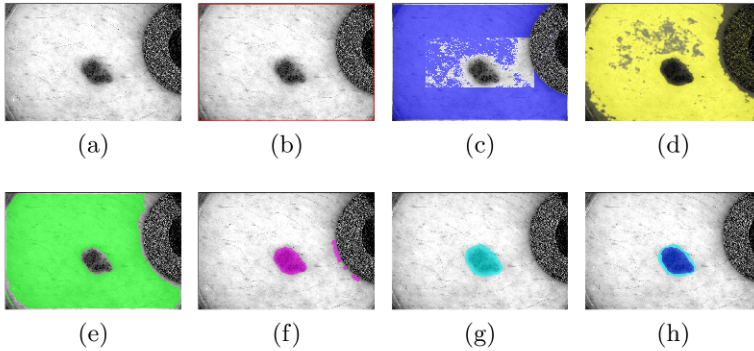


Fig. 3. Segmentation of image ISIC_0008294. Figure (a) shows the nevus intensities (greyscale); (b) shows the border of the image (red); (c) shows the sample of the skin (blue); (d) shows the similarity score of each pixels w.r.t. the sample of the skin (variation of the score is shown as the varying 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). (Color figure online)

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*⁸. Programming languages are used in conjunction with libraries of imaging primitives⁹, and with Machine Learning (ML) libraries, such as *Keras*¹⁰.

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.

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. [20]). These usually consist of a series of steps and checks that the Radiotherapist

⁸ See <https://www.python.org/>.

⁹ 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/>.

¹⁰ See <https://keras.io/>.

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 [43], 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. Spatial logics instead are very close to the domain of discourse, and permit rapid specification of properties of interest that are related to the spatial distribution of imaging features. Also note that it is not guaranteed that a hand-written 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*[42], automatic parallel execution (see [9]), on-disk caching, and *graphical processing units* (see [11]), are used to speed up computation.

Example 4. Similarly to Example 3, but in a different application context. 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. More precisely, first, ML-based annotators would be executed on an imaging data set. Then, a `VoxLogicA` specification can be written that loads the ML-based annotations, and combines them using logical operators.

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. Each centre could contribute a portion of a `VoxLogicA` specification and practitioners would be entitled to view both the specification and the results in order to quickly gather as much information as available on each patient. As an example, consider queries related to tumour size, past growth, proximity to organs at risk, etc.

6 Outlook: `VoxLogicA` as a Distributed Query Engine

In the opinion of the authors, the tool `VoxLogicA` as is, would already constitute a good starting point for querying medical imaging datasets. This is because `ImgQL` is endowed with *global* operators, e.g. for computing the number of voxels that evaluate to true, that is the *volume*, of a given formula, or for comparing different regions to check if they are included in one another, or are overlapped, or are disconnected etc.¹¹ By this, using `VoxLogicA` alone, it is possible to identify in a dataset of medical images those satisfying certain global properties, based on the segmentation provided by the voxel-based properties (such as those described in Sect. 4). For a concrete example, consider e.g. the ability to find all the images of patients where *the brain tumour is larger than x* , or *the oedema is larger than the tumour*, or *the gross tumour volume is very close to the cerebellum*, or, changing the example, where a nevus is *divided into k regions*, and so on. For the nevi segmentation case study, also properties related to the shape and contour of a nevus are very relevant. We also emphasize that the “lower level” information, e.g. the tumour segmentation, or the skull segmentation, could also be the output of Machine-Learning based voxel classifiers, but still, reachability, proximity, overlap, volume of the regions obtained

¹¹ Such operators come from the classical theory of *region calculi*, see [16] for details.

via Machine Learning would be the ingredients of complex queries that could be run using our tool. We note in passing that the current language is textual, although future work could also consider its *visual* counterpart¹². The reader should keep in mind that **VoxLogicA** is aimed at a technical audience, namely that of *domain experts* in medical imaging, who generally speaking are engineers, physicists, MDs or technicians, with few programming expertise, but quite a lot of technical competence.

However, a number of research directions could be pursued, in order to turn the current tool into a fully-fledged query engine. In this section we sketch a research roadmap, by describing a number of future improvements, to direct our results more closely to the field of Information Retrieval.

Integration with Clinical Databases. Information systems such as *RIS-PACS* [26] 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 [17, 18], 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 [38]). Also for multi-centric studies, in order to respect privacy and intellectual property concerns, it would be worthwhile 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 OpenCL¹³, an

¹² It is out of the scope of this work to discuss the advantages and disadvantages of visual vs. textual query languages.

¹³ See <https://www.khronos.org/opencl/>.

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 prototype 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 dataset, 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 [29], and the cognitive load issues investigated in [10]).

7 Concluding Remarks

In this paper we have sketched our research roadmap to turn the spatial model checker **VoxLogicA** into a fully-fledged query engine for datasets of medical images. Some of these directions will already be pursued in the immediate future, namely the on-GPU implementation, the Human-computer Interaction aspects, and the minimization procedures. At the horizon, distributed execution, and integration with clinical databases, could be the last steps needed, in order to initiate the technological transfer of our results in this direction.

References

1. XQuery 3.1: An XML Query Language (2017). <https://www.w3.org/TR/2017/REC-xquery-31-20170321/>
2. Aiello, M., Pratt-Hartmann, I., Benthem, van, J.: Handbook of Spatial Logics. Springer, Heidelberg (2007). <https://doi.org/10.1007/978-1-4020-5587-4>
3. Ali, A.R.A., Deserno, T.M.: A systematic review of automated melanoma detection in dermoscopic 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)
4. Aswani, N., 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)

5. 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 (2019). <https://doi.org/10.1007/s10009-019-00511-9>
6. Bartocci, E., Bortolussi, L., Loretì, M., Nenzi, L., Silvetti, S.: MoonLight: a lightweight tool for monitoring spatio-temporal properties. In: Deshmukh, J., Ničković, D. (eds.) *RV 2020. LNCS*, vol. 12399, pp. 417–428. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-60508-7_23
7. Belmonte, G., Broccia, G., Vincenzo, C., Latella, D., Massink, M.: Feasibility of spatial model checking for nevus segmentation. In: *Proceedings of the 9th International Conference on Formal Methods in Software Engineering (FormalISE 2021)*, pp. 1–12. IEEE (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: *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)
9. Belmonte, G., Ciancia, V., Latella, D., Massink, M.: *VoxLogicA*: a spatial model checker for declarative image analysis. In: Vojnar, T., Zhang, L. (eds.) *TACAS 2019. LNCS*, vol. 11427, pp. 281–298. Springer, Cham (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)
11. Bussi, L., Ciancia, V., Gadducci, F.: Towards a spatial model checker on GPU. In: Peters, K., Willemse, T.A.C. (eds.) *FORTE 2021. LNCS*, vol. 12719, pp. 188–196. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78089-0_12
12. Chamberlin, D.D., Boyce, R.F.: SEQUEL: a structured English query language. In: Rustin, R. (ed.) *Proceedings of 1974 ACM-SIGMOD Workshop on Data Description, Access and Control*, pp. 249–264. ACM (1974)
13. Ciancia, V., Gilmore, S., Grilletti, G., Latella, D., Loretì, M., Massink, M.: Spatio-temporal model checking of vehicular movement in public transport systems. *Int. J. Softw. Tools Technol. Transf.* **20**(3), 289–311 (2018). <https://doi.org/10.1007/s10009-018-0483-8>
14. Ciancia, V., Latella, D., Loretì, M., Massink, M.: Specifying and verifying properties of space. In: Diaz, J., Lanese, I., Sangiorgi, D. (eds.) *TCS 2014. LNCS*, vol. 8705, pp. 222–235. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44602-7_18
15. Ciancia, V., Latella, D., Loretì, M., Massink, M.: Model checking spatial logics for closure spaces. *Log. Methods Comput. Sci.* **12**(4) (2016)
16. Ciancia, V., Latella, D., Massink, M.: Embedding RCC8D in the collective spatial logic CSLCS. In: Boreale, M., Corradini, F., Loretì, M., Pugliese, R. (eds.) *Models, Languages, and Tools for Concurrent and Distributed Programming. LNCS*, vol. 11665, pp. 260–277. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21485-2_15
17. Ciancia, V., Latella, D., Massink, M., de Vink, E.: Towards spatial bisimilarity for closure models: logical and coalgebraic characterisations (2020). <https://arxiv.org/abs/2005.05578>
18. Ciancia, V., Latella, D., de Vink, M.M.E.: On bisimilarities for closure spaces - preliminary version (2021). <https://arxiv.org/abs/2105.06690>
19. Codella, N.C., et al.: Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM J. Res. Dev.* **61**(4/5), 5–1 (2017). Special Issue on Deep Learning

20. Combs, S.E., et al.: ESTRO ACROP guideline for target volume delineation of skull base tumors. *Radiother. Oncol.* **156**, 80–94 (2021)
21. Demner-Fushman, D., Müller, H., Kalpathy-Cramer, J., Antani, S.: A decade of community-wide efforts in advancing medical images understanding and retrieval. *Computer. Med. Imaging Graph.* (2014)
22. Deserno, T.M., Güld, M.O., Thies, C., Plodowski, B., Keysers, D., Ott, B., Schubert, H.: IRMA - content-based image retrieval in medical applications. *Stud. Health Technol. Inform.* **107**(Pt 2), 842–6 (2004)
23. Galton, A.: A generalized topological view of motion in discrete space. *Theor. Comput. Sci.* **305**(1–3), 111–134 (2003)
24. Godar, D.E.: Worldwide increasing incidences of cutaneous malignant melanoma. *J. Skin Cancer* 2011 (2011)
25. 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)
26. Haux, R.: *Strategic Information Management in Hospitals: An Introduction to Hospital Information Systems*. Springer, New York (2003). <https://doi.org/10.1007/978-1-4757-4298-5>
27. 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. *Int. J. Healthc. Inf. Syst. Inform. Off. Publ. Inf. Resour. Manag. Assoc.* **4**(1), 1–16 (2009)
28. Lozano, R., 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. *Lancet* **380**(9859), 2095–2128 (2012)
29. 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. Med. Inform.* **84**(10), 774–783 (2015)
30. Mazzara, G., Velthuisen, R., Pearlman, J., Greenberg, H., Wagner, H.: Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. *Int. J. Radiat. Oncol. Biol. Phys.* **59**(1), 300–312 (2004)
31. Menze, B.H.e.a.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**(10), 1993–2024 (2015)
32. Mocellin, S., Nitti, D.: Cutaneous melanoma in situ: translational evidence from a large population-based study. *Oncologist* **16**(6), 896 (2011)
33. 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, MULTIMEDIA 2005*, pp. 1014–1022 Association for Computing Machinery, New York (2005)
34. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *Int. J. Med. Inform.* **73**(1), 1–23 (2004)
35. Nenzi, L., Bortolussi, L., Ciancia, V., Loreti, M., Massink, M.: Qualitative and quantitative monitoring of spatio-temporal properties with SSTL. *Log. Methods Comput. Sci.* **14**(4), 1–38 (2018)
36. Randell, D.A., Cui, Z., Cohn, A.G.: A spatial logic based on regions and connection. In: *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning (KR 1992)*, pp. 165–176. Morgan Kaufmann (1992)
37. Randell, D.A., Galton, A., Fouad, S., Mehanna, H., Landini, G.: Mereotopological correction of segmentation errors in histological imaging. *J. Imaging* **3**(4) (2017)

38. Spyridon (Spyros) Bakas et al. (Ed.): 2017 International MICCAI BraTS Challenge: pre-conference proceedings, September 2017. https://www.cbica.upenn.edu/sbia/Spyridon.Bakas/MICCAI.BraTS/MICCAI_BraTS_2017_proceedings_shortPapers.pdf
39. Tang, L.H., Hanka, R., Ip, H.H.: A review of intelligent content-based indexing and browsing of medical images. *Health Inform. J.* **5**(1), 40–49 (1999)
40. 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 (2017)
41. Vorwerk, H., et al.: 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 (2014)
42. Wikipedia contributors: Memoization – Wikipedia, the free encyclopedia. <https://en.wikipedia.org/w/index.php?title=Memoization&oldid=1002117560> (2021). Accessed 29 Mar 2021
43. 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*. LNCS, vol. 8442, pp. 611–626. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-06410-9_41
44. Zong, H., Verhaak, R.G., Canoll, P.: The cellular origin for malignant glioma and prospects for clinical advancements. *Expert Rev. Mol. Diagn.* **12**(4), 383–394 (2012)