Topological Radiomics (TOPiomics): Early Detection of Genetic Abnormalities in Cancer Treatment Evolution

Debora; Gil,[1*] Arnau; Mas[1] , Oriol; Ramos[1] , Raquel; Pérez[2]

[1]Computer Vision Center; edifici O, campus UAB, 08193, Bellaterra (Barcelona), Spain [2] Vall d'Hebron Institute of Oncology; Paseo Vall d'Hebron, 119 - Ed Materno Infantil, 08035, Barcelona
*debora@cvc.uab.es

ABSTRACT
A main challenge during anti-cancer treatment is early detection of variations in tumour phenotype altering the expected outcome. Current evaluation using biopsy is not appropriate since samples may not be representative of the whole lesion. Radiomics is an emerging area that converts medical imaging data into measures (imaging phenotype) of the whole tumour correlated with genomics.

TOPiomics is the 1st imaging technique for early detection of variations in tumour imaging phenotype altering response of anti-cancer treatments. TOPiomics topological description can model the complex structure of radiomics data and detect abnormal imaging phenotypes related to outcome in cancer treatment.

Keywords: Personalized anti-cancer treatment; radiomics; outlier detection

1. INTRODUCTION

In the era of precision medicine, cancer therapies are tailored to the specific genetic makeup of a tumour. A main challenge during treatment is the early detection of variations in tumour phenotype that might alter the expected outcome. Current evaluation of genomic aberrations using biopsy is not appropriate for treatment follow-up, since procedures are invasive, cannot be repeated frequently and samples may not be representative of the whole lesion.

Under the rational that radiomics signatures correlate with tumour phenotypes and genomic aberrations in cancer, abnormalities in radiomic features can be potentially used as predictive and early response biomarkers to anti-cancer treatments. Detection of abnormal radiomic features should model multi-view spaces with Small Sample Size (SSS) data prone to have a complex manifold structure in radiomics high dimensional spaces. TOPiomics is the 1st imaging technique specific for early detection of variations in tumour imaging phenotype altering the response of anti-cancer treatments. TOPiomics uses algebraic topology to model the complex structure of radiomics SSS multi-view data ensuring reproducible clinical results. Its non-parametric local description endows TOPiomics with high robustness to detect abnormalities in SSS contexts, while its view-sensitive approach allows early detection of abnormal imaging phenotypes. Therefore, TOPiomics is a unique specific technique to define robust imaging biomarkers for outcome in cancer treatment follow-up that will improve cancer patients care by optimizing treatment selection and sequence.

TOPiomics has obtained the results in the areas of machine learning problems and personalized medicine (pending final clinical interpretation and validation). The results related to machine learning problems are the following:

- Local description of feature spaces able to model complex distributions with high performance in SSS unbalanced sets.
- Modelling of intrinsic (topological) properties of heterogeneous multi-view feature spaces.
- Topologically invariant scores (persistent homology) of datasets structure that allow matching of multi-view data.
- Abnormality normalized representation space with universal measure applicable to any data without training.

The (expected) results related to personalized medicine are the following:

- Relation between radiomics and genomic features to improve the clinical interpretation of radiomic signatures.
- (Pending) Detection of non-responders to anti-cancer treatments.

2. STATE OF THE ART

Current evaluation of genomic aberrations relies on acquisition of tumour samples from biopsies. Since biopsies are invasive, it is not possible to repeat them

several times along treatment. Besides, biopsies represent only a small portion of the tumours, which may not be representative of the whole lesion. The most widely used medical imaging modality in oncology is computed tomography (CT), which assesses tissue density. Due to the non-invasive nature of CT, its capacity of study the whole body and its broad use in clinical practice, CT-radiogenomics offers enormous promise for novel imaging biomarkers development.

Currently there are no methods for early detection of variations in anticancer outcome from abnormalities in tumour imaging phenotype. Since early detection of variations in tumour imaging phenotype bases on abnormal values in radiomic features, the closest techniques are methods for abnormality detection developed in the context of outlier detection in statistical models and model training for classification. Existing methods for detection of outliers can be categorized into global (including deep methods) and local approaches.

Global methods are population based and model the distribution in the feature space of a set of (annotated) samples using parametric descriptions. In SSS unbalanced problems like radiomics, data distribution cannot be properly estimated which leads to lack of reproducibility (over fitting). Also abnormalities become influential points that bias global approaches and drop their potential to detect abnormal cases.

Local methods are based on a description of the structure of each sample's neighbours in the feature space. Although they are better suited in SSS, current approaches based on Euclidean metrics are unable to describe complex multi-view data. Also, the selection of the parameters defining neighbourhoods is a main bottleneck. They depend on dimensionality, no criteria to set them exists, so they are heuristically tuned for each dataset.

3. BREAKTHROUGH CHARACTER OF THE PROJECT

TOPiomics is **the first imaging technique** specific for **early detection of variations** in tumour imaging phenotype **altering the response to anti-cancer treatments**. Unlike existing methods for detection of abnormalities in feature spaces, TOPiomics bases on a topological description of radiomics multi-view data. Topology is a powerful mathematical approach to model the structure of complex manifolds without the assumption of any parametric model for the data. As a consequence, TOPiomics is a flexible technique able to model the complex structure of radiomics SSS data without overfitting, thus, ensuring reproducible clinical results. Its non-parametric local description endows it with high robustness to detect abnormalities in SSS contexts, while its view-sensitive approach allows early detection of abnormal imaging phenotypes. Thus, TOPiomics is a unique specific technique for defining robust imaging biomarkers for outcome in cancer treatment follow-up that will improve cancer patients care by optimizing treatment selection and sequence.

Comparing to state of the art outliers detection methods, TOPiomics has several advantages. On one side, our description of feature spaces local structure endows TOPiomics with the following properties:

- Capability to model complex data distributions structured as manifolds. Our description based on topology provides, for each sample, with a set of multiple neighbourhoods alternative to the usual n-dimensional Euclidean balls centred at each sample. TOPiomics alternative neighbourhoods can describe multi-view data distributions structured as manifolds, which is a significant advantage over existing local methods.
- High performance in SSS unbalanced sets. Unlike global and deep learning methods, topology can be robustly computed for SSS regardless of the dimension of feature spaces. This is a must for the assessment of experimental treatments and early diagnosis of rare cases.

On the other side, the normalized representation of outlierness has the following significance:

- Ability to detect outliers in heterogeneous multi-view settings. Existing methods cannot properly match data across multi-view representations, like radiomic and genomic. TOPiomics invariant scores (persistent homology) of datasets structure allow matching heterogeneous multi-view data to facilitate better clinical interpretation of radiogenomic data.
- Universal outlierness measure applicable to any data without training. TOPiomics only requires the computation of persistent homology of each new data set. This is a main advantage over most methods that require training the model from scratch for each new data set and endows TOPiomics with high scalability to any pathology and treatment.

Table 1 summarizes TOPiomics breakthrough character.

Tab. 1. TOPiomics breakthrough for detection of alterations in anti-cancer treatments outcome.

	GLOBAL	LOCAL	TOPIOMICS
Low-cost training (high scalability to any pathology)	NO	YES	YES

Model complex data (model radiogenomic spaces)	NO	PARTIAL	YES
Model heterogeneous multi-view data (Relation between radiomic and genomic data)	NO	NO	YES
High performance in SSS (detection of rare cases/mutations; assessment of experimental treatments and response in new pathologies)	NO	PARTIAL	YES

4. PROJECT RESULTS

TOPiomics aims at defining a measure of abnormality in tumour imaging phenotype altering response to anti-cancer treatments using a topological description of radiomic spaces. The topological description is based on the structure of the mutual k-nearest neighbour graph, MKG (Fig. 1 (b)). Such structure has been modelled using 2 approaches: 1) graph communities [1]; 2) graph persistent homology [2].

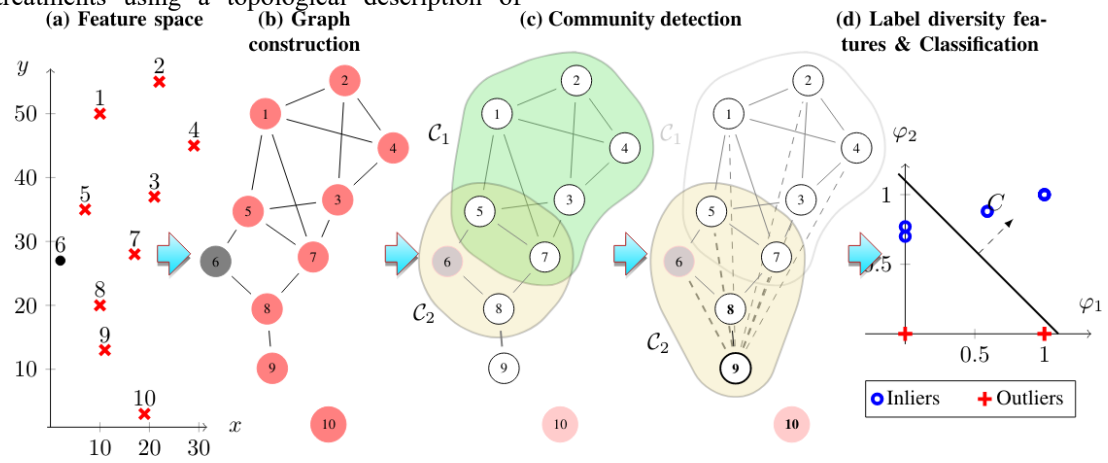

Fig. 1. Graph Communities: (a) Feature space with 2 classes (black dots and red crosses). (b) MKG with nodes coloured according to its class. Nodes 6 and 10 are, respectively, a class (samples labelled differently from neighbours) and attribute (isolated samples with abnormal features) outliers. (c) Communities: initial in left graph and their extension in the right one. (d) Outlierness space

Detection based on Graph Communities

Graph communities (Fig. 1 (c)) are clusters of nodes with specific connectivity properties. We use criteria for dynamic computation of communities to extend an initial set of communities given by Percolation clusters (left graph in Fig 1 (c)) which are defined as maximal unions of adjacent k-cliques. Since Percolation communities are prone to exclude points that are not outliers, we extend them (right graph of Fig 1 (c)) using the community internal connectivity and the connections between new nodes and community nodes.

Nodes not belonging to any community correspond to attribute outliers (isolated samples with abnormal features). Meanwhile, class outliers belong to communities with high heterogeneity in nodes label. We have defined a measure φ1 quantifying the heterogeneity in community labels and a φ2 quantifying how many nodes in the community have a label different from the sample label. Measures are normalized in [0,1] to define a function, φ=(φ1, φ2), mapping inliers, attribute and class outliers to different corners of the square [0,1]x[0,1] (Fig 1 (d)). The probability of a classifier trained to discriminate between them provides our outlierness score (Fig 1 (d)).

We have assessed this approach in UCI datasets altered to have attribute and class outliers in single and multi-view settings and including SSS unbalanced classes. TOPiomics outperforms existing state-of-art methods in, both, single and multi-view settings, regardless of the outlier configuration and performing equally well in SSS unbalanced datasets. Although not critical, one weakness is the choice of k for the computation of MKG. This motivates the second implementation.

Detection based on Graph Persistent Homology

Topological spaces can be assembled out of glued together simplices. Simplices are generalisations of points, lines, triangles and tetrahedra to higher dimensions and for graphs they correspond to cliques. Unions of simplices that are not the boundary of a higher dimensional simplex define a “hole” of the topological space since they do not enclose any part of it.Two topological spaces having the same structure of holes can be considered equivalent. Homology groups algebraic tools that encode k-dimensional holes and, thus, can be used to match topological spaces. Since the 0 homology group encodes the number of connected components of the space, it can be used to detect outliers and rare cases as connected components of small size.

To alleviate dependence on the parameter k, persistent homology computes homology for every value to determine which features really reflect geometric aspects. Some features might be prominent only at specific k, whether others will persist for a large range of k. If a point is an outlier, its homology class 0 should live for a considerable time and have a small number of points. Thus, the normalised lifetime and the % of points of each 0 homology class detect outliers as points at the lower right-hand corner (Fig. 2).

We conducted a 1st assessment in 20 patients to validate correlation between radiomic and genomic (RNA) alterations. Radiomics were given by 26 PyRadiomic features selected by VHIO. Both spaces (Fig. 3) have cluster 12 with similar persistence and patient 6 is a genomic outlier.

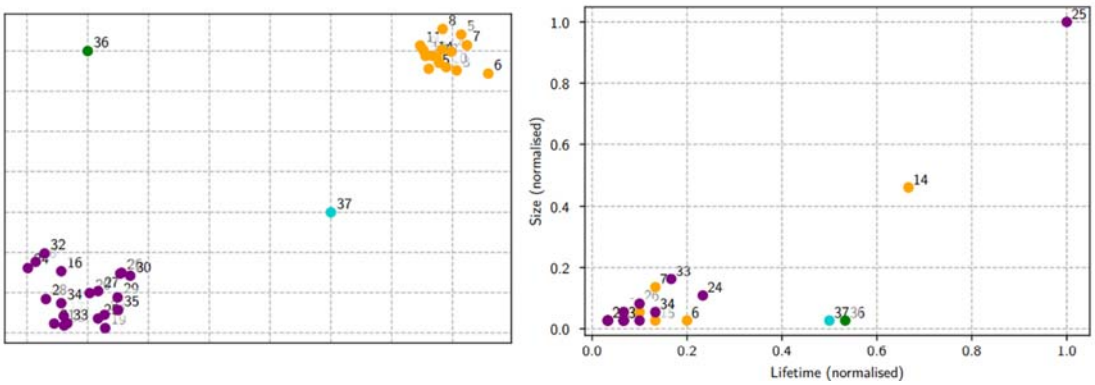

Fig. 2. Persistent Homology. In top, a custom data set with two clusters and two outliers and in bottom persistent homology space.

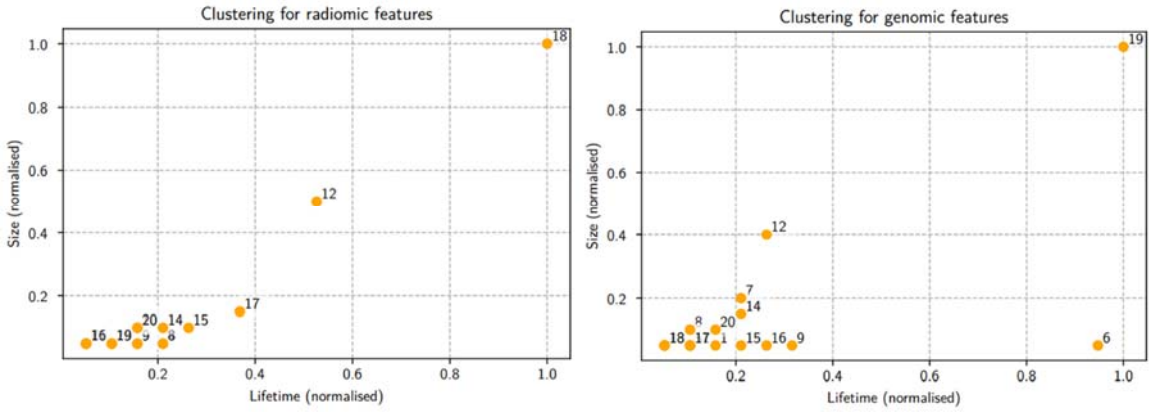

Fig. 3. Radiogenomic persistent space with points labelled with patient number.

5. FUTURE PROJECT VISION

5.1. Technology Scaling

In Phase 1, we have carried out a proof of concept with radiogenomic data of 20 patients with personalized anti-cancer treatment. In order to reach TRL 5-7, TOPiomics should be deployed into clinical institutions to be tested in other sets of patients for a multi centric trial. With the aim of assessing the real market need as diagnosis and treatments follow-up tool, the technology will be tested in other pathologies with experimental personalized treatments requiring follow-up.

To facilitate wide deployment in clinical centers, TOPiomics will be implemented in a heterogeneous platform as a Software as a Service (SaaS) diagnosis tool to provide a cost-effective solution to clinicians and other potential users. The core computation of the persistent homology based on cliques will be accelerated using parallel implementation in GPU or FPGA.

5.2. Project Synergies and Outreach

In Phase 2, we will incorporate:

- Technological partner (like CEPHIS-UAB) able to provide cloud services and high performance computing to implement TOPiomics in a heterogeneous platform deployed as SaaS.
- Technological partner (like Galgo or Crisalix) to manage access to PACs data base to

download anonymized clinical data and send it to cloud services.

- Clinical centers to provide data from pathologies requiring identification of non-responders to validate TOPiomics actual value. We will use our clinical partners in Spain (Hospital Bellvitge, Hospital de Can Ruti, Hospital del Mar) and their international collaborations (e.g. UZ Leuven, Policlinico Sant'Orsola).

We will use the MarketPlace platform provided by the ATTRACT Consortium and the 1:1 meetings to be held at ATTRACT final conference to identify ATTRACT Phase 1 funded projects (like MERIT-VA or PIZZICATO) to cluster with.

Diffusion will target the clinical community, commercial companies and public at large. We plan to regularly update the project website, implement a demo showcasing the capabilities of each prototype for face to face demonstrations in medical events (like MEDICA) and trade shows (like Mobile World Congress), create social media accounts, news in press and mass-media, record promotional videos and publish scientific papers.

5.3. Technology application and demonstration cases

TOPiomics benefits the area of health, demographic change and wellbeing and in Phase 2 we plan to implement the following demonstration cases:

- Optimize treatment selection and development of experimental treatments. Early detection of abnormal cases allows assessing the efficiency of new treatments in oncology and emerging pathologies (like COVID-19).
- Optimize biopsy by the detection of unknown/rare mutations in medical scans. This will increase efficiency of screening programs by better detecting rare cases having a higher probability of developing cancer.
- Fuse radiological and clinical data for better prediction of rare/unknown pathologies like COVID-19 or genetic.

Regarding RI communities, we could contribute to clinical data repositories like https://ecrin.org/.

5.4. Technology commercialization

We plan to do an IP Protection study to validate the most suitable strategy for technology protection. The reports generated will help determining different transfer options to the medical industry.

We plan to commercialise TOPiomics as a SaaS in a heterogeneous platform and also to study the feasibility of a local deployment. As we have done for other projects, we will join acceleration programs (like Caixaimpulse, Genesis Biomed, The Collider, EITHealth) with the aim to raise awareness to different venture capitals and future entrepreneurs. We will also use digital platforms like Ekiter to validate the constitution of a deep tech company.

5.5. Envisioned risks

Table 2 reports risks, impact and mitigation actions.

Tab. 2. Risks.

RISK	IMPACT	MITIGATION
Failure to detect alterations with current radiomic features	Medium	Enlarge radiomic space adding other Pyradiomics feeatures and deep
No relation between radiomics and genomics	Low	Consider more measures, like distance between homology groups, diameter and volume

5.6. Liaison with Student Teams and Socio-Economic Study

During Phase 1, we liaised with an undergraduate student of the degree in Mathematics and Physics in his last course (equivalent to Msc Level) through a CVC program. His role was to advise on how to implement persistent homology in graphs.

For Phase 2 we plan liaison with Msc students in science and entrepreneurship from ESADE and EADA through the programs MAP-EADA and EMPENTA-ESADE to carry on a socio-economic study like the ones carried out (*https://www.questionpro.com/t/AQ9OCZieca*) for other projects We will also engage Fusion Point, initiative designed to find innovative solutions to real-life challenges through interdisciplinary work.

6. ACKNOWLEDGEMENT

This project has received funding from the ATTRACT project funded by the EC under Grant Agreement 777222, Spanish projects FIS-G64384969, RTI2018-095209-B-C21, Generalitat de Catalunya 2017-SGR-1624 and CERCA. DGil is Serra Hunter Fellow. R.Perez is supported by PCF-Young Investigator Award and PI18/01395

7. REFERENCES

[1] O. Ramos Terrades, A. Berenguel, D. Gil., 2020. A flexible outlier detector based on a topology given by graph communities, https://arxiv.org/abs/2002.07791.

[2] Mas, A., 2020, Topological data analysis and persistent homology. Developing an outlier detector based on persistent homology, Mathematics BSc. Undergraduate Thesis, UAB.