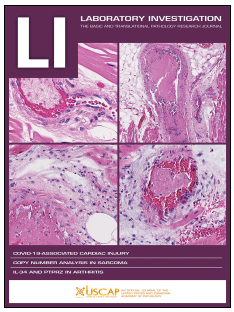


Advancing Precision Medicine: Algebraic Topology and Differential Geometry in Radiology and Computational Pathology

Richard M. Levenson, Yashbir Singh, Bastian Rieck, Quincy A. Hathaway, Colleen Farrelly, Jennifer Rozenblit, Prateek Prasanna, Bradley Erickson, Ashok Choudhary, Gunnar Carlsson, Deepa Deepa

PII: S0023-6837(24)01738-0

DOI: https://doi.org/10.1016/j.labinv.2024.102060

Reference: LABINV 102060

To appear in: *Laboratory Investigation*

Received Date: 15 December 2023

Revised Date: 8 April 2024

Accepted Date: 10 April 2024

Please cite this article as: Levenson RM, Singh Y, Rieck B, Hathaway QA, Farrelly C, Rozenblit J, Prasanna P, Erickson B, Choudhary A, Carlsson G, Deepa D, Advancing Precision Medicine: Algebraic Topology and Differential Geometry in Radiology and Computational Pathology, *Laboratory Investigation* (2024), doi: https://doi.org/10.1016/j.labinv.2024.102060.

Advancing Precision Medicine: Algebraic Topology and Differential Geometry in Radiology and Computational Pathology

Richard M Levenson[1*], Yashbir Singh[2*], Bastian Rieck[3], Quincy A. Hathaway[4], Colleen Farrelly[5], Jennifer Rozenblit[6], Prateek Prasanna[7], Bradley Erickson[2], Ashok Choudhary[8], Gunnar Carlsson[9], Deepa Deepa[10]

Affiliations:
1. Department of Pathology and Laboratory Medicine, UC Davis, California, USA
2. Department of Radiology, Mayo Clinic, Rochester, MN, USA
3. Helmholtz Munich and Technical University of Munich, Munich, Germany
4. Department of Medical Education, West Virginia University, Morgantown, WV, USA
5. Post Urban Ventures, London, United Kingdom
6. Department of Mathematics, The University of Texas, Austin
7. Department of Biomedical Informatics, Stony Brook, NY, USA
8. Department of Surgery, Mayo Clinic, Rochester, MN, USA
9. Department of Mathematics, Stanford University, California, USA
10. Institute of Genomic Health, Ichan school of Medicine, Mount Sinai, NY, USA

*Corresponding author and Sharing First authorship

Professor Richard M. Levenson, M.D.
Professor and Vice Chair for Strategic Technologies,
Department of Pathology and Laboratory Medicine, Pathology Building
4400 V St.
Sacramento, UC Davis, CA 95817

Yashbir Singh, M.Tech, PhD
Assistant Professor, Radiology, Mayo clinic, Rochester, MN
200 1st St SW, Rochester, MN 55905

Abstract

Precision medicine aims to provide personalized care based on individual patient characteristics, rather than guideline-directed therapies for groups of diseases or patient demographics. Images—both radiology- and pathology-derived—are a major source of information on presence, type, and status of disease. Exploring the mathematical relationship of pixels in medical imaging ("radiomics") and cellular-scale structures in digital pathology slides ("pathomics") offers powerful tools for extracting both qualitative, and increasingly, quantitative data. These analytical approaches, however, may be significantly enhanced by applying additional methods arising from fields of mathematics such as differential geometry and algebraic topology that remain underexplored in this context.

Geometry's strength lies in its ability to provide precise local measurements, such as curvature, that can be crucial for identifying abnormalities at multiple spatial levels. These measurements can augment the quantitative features extracted in conventional radiomics, leading to more nuanced diagnostics. By contrast, topology serves as a robust shape descriptor, capturing essential features such as connected components and holes. The field of topological data analysis was initially founded to explore the shape of data, with functional network connectivity in the brain being a prominent example. Increasingly, its tools are now being used to explore organizational patterns of physical structures in medical images and digitized pathology slides. By leveraging tools from both differential geometry and algebraic topology, researchers and clinicians may be able obtain a more comprehensive, multi-layered understanding of medical images and contribute to precision medicine's armamentarium.

Keywords: Precision medicine, pathomics, radiomics, topological data analysis, geometry

Introduction

In the era of precision medicine, medical researchers and practitioners continue to seek improvements in diagnostic accuracy, prognosis prediction, and treatment selection. Image-based tools that allow caregivers to visualize and interpret anatomical structures and identify abnormalities represent a major arena in which such advances are being explored. Recent developments in the field of radiomics and pathomics can transform 2- and 3-dimensional images into elaborate matrices of data and information that can generate insights with potential clinical benefit [1-3]. Radiomics is a rapidly evolving discipline that focuses on the extraction and analysis of quantitative features from medical images, employing algorithms to extract a wide range of quantitative features including shape, intensity, texture, and spatial patterns. These features can provide valuable information about tumor heterogeneity, treatment response and patient outcome. By utilizing machine learning and artificial intelligence techniques radiomics may allow development of novel predictive models and personalized medicine approaches [4,5]. Pathomics, similarly, delves into the analysis of digitized histopathology slides. Histopathology has long been considered the gold standard for cancer diagnosis and grading, but the manual interpretation of histological images is both subjective and time-consuming. Pathomics techniques aim to overcome these limitations by leveraging computational techniques to analyze digitized slides and extract quantitative information related to tissue morphology, cellular architecture, and molecular markers. By quantifying histological features, pathomics can similarly enable objective assessment while providing insights into disease progression, response to treatment, and patient prognosis [6-8]. Combining information from multiple modalities allows for a more comprehensive understanding of diseases and the integration of radiomics and pathomics into decision-making workflows has the potential to advance precision medicine by bridging the gap between imaging, pathology, and clinical data. Furthermore, radiomics and pathomics offer the opportunity to uncover previously hidden patterns and biomarkers that may have significant clinical implications, paving the way for personalized therapeutic interventions and improved patient outcomes [3,9-12].

In this overview, we will explore how algebraic topology and differential geometry can augment the fields of radiomics, pathomics, and multiomics-based techniques, delving into the various applications and challenges associated with these emerging disciplines, including new tools from geometry and topology to improve model accuracy and integration with other data sources. We will discuss the potential benefits and examine the current state of research and clinical implementation. In particular, with respect to machine (or deep) learning approaches methods being developed and deployed [6,13,14], it appears that prior extraction of topological and geometric features can help improve both training times and performance in certain cases.

Overview of Precision Medicine

Precision medicine, also known as personalized medicine or stratified medicine, is an evolving approach to healthcare that aims to provide tailored medical interventions to individual patients based on their unique characteristics. Unlike the traditional one-size-fits-all approach, precision medicine recognizes that each patient's genetic makeup, lifestyle

factors, environmental influences, and disease characteristics are distinct and should be taken into account when making diagnostic and treatment decisions [15-18]. This comprehensive approach ideally allows healthcare providers to develop a holistic understanding of a patient's health status and disease progression. Despite its potential, precision medicine faces several challenges. The interpretation and integration of vast amounts of data, the need for robust analytical tools, ethical considerations regarding data privacy and consent, and equitable access to personalized treatments are only some of the hurdles that need to be addressed. However, in the meantime, it is becoming practical to combine insights from closely related arenas, such as radiomics and pathomics; combining quantitative features from both medical images and histopathology slides is already providing valuable insights into tumor characteristics, treatment response, and prognosis [2,19,20].

Some FDA-approved, mathematically sophisticated tools are already available. 3D Slicer, an early example of a biomedical image analysis platform [21] demonstrates the integration of mathematical principles into radiology. PathAI [22] employs deep learning and AI to assist pathologists in diagnosing diseases more accurately, highlighting the growing influence of mathematical algorithms in pathology. TexRAD [23] is an FDA-cleared radiomics software that utilizes texture analysis to extract intricate patterns from PET scans, providing quantitative insights into tissue heterogeneity. Texture analytics are also promising, especially when combined with multimodal information [24]. Many tools are also emerging in the AI-pathology space, and their development may democratize the availability of high-quality healthcare analytics and diagnostics, bridging gaps in global health disparities while maintaining privacy and data protection.

Radiomics

Radiomics pipelines focus on extracting and analyzing quantitative features from medical images and have recently demonstrated promise across several areas, including tumor characterization, treatment response assessment, prognostic prediction, and treatment planning [9,25]. By combining radiomics features with clinical data including as patient demographics, laboratory values, and histopathological information, these models can be further enriched. Radiomics has also shown promise in treatment response assessment by analyzing changes in quantitative features over time as part of monitoring treatment effectiveness. For example, in the context of cancer patients undergoing chemotherapy, radiomics analysis of serial imaging scans plays a crucial role in assessing treatment response, allowing for timely adjustments to the treatment regimen [26]. Additionally, if a treatment regimen can be initiated earlier, this can impact prognosis. Traditional radiomics features, such as first-order statistics, shape-based features, and texture-based features, are employed to quantify various aspects of the medical images [3,5,7,9,26]. While radiomics has made significant strides, several challenges remain, including a lack of topology and geometry-based analytics. Radiomics feature extraction and analysis methods can vary across studies, leading to inconsistent results and limited comparability. Efforts are underway to establish standardized protocols and feature sets to improve reproducibility and facilitate multicenter collaborations [3,27].

Pathomics

Pathomics focuses on the analysis of digitized histopathology slides, and more recently, on developing methods for slide-free histology in digital pathology. With the increasing availability of whole-slide imaging scanners and advancements in computational pathology, pathomics has gained momentum in recent years. Pathomics allows for the extraction of quantitative histology-based information, enabling objective assessment of tissue morphology, cellular characteristics, and molecular markers, aiding in the classification and grading of tumors. By quantifying histological features from tissue samples using image processing techniques, machine learning algorithms, and deep learning models, pathomics can assist in identifying patients who may benefit from specific treatments or who are at higher risk of disease recurrence [9]. Pathomics has been used to identify lung cancer subtypes [28,29], lung adenocarcinoma [30], and, combined with a variety of genomic techniques such as transcriptomics [31], represents an opportunity for highly multiplexed (tens to thousands of analytes), high spatial resolution ("spatial omics") data to be investigated [32,33]. Despite its potential, pathomics faces challenges like those confronting radiomics, including standardization, reproducibility, and scalability. The digitization of histopathology slides, data storage and management, and the integration of pathomics with other data modalities pose technical and logistical hurdles. Moreover, the transition from manual pathology evaluation to computational analysis requires rigorous validation and regulatory acceptance.

Algebraic Topology and Differential Geometry

Advanced mathematical concepts from algebraic topology and differential geometry are still being explored within healthcare; these are two branches of mathematics with diverse applications. Topology aims to uncover characteristic properties—invariants—of a space (or data when connections and relationships can be defined spatially). These invariants serve to characterize a space up to certain transformations: for instance, the number of points in a dataset is a simple invariant that remains unchanged under rotation. One of the more involved but also more powerful invariants is homology, a concept central to algebraic topology. It permits making statements about connectivity characteristics of a space by means of algebraic calculations. Tools commonly applied in data analysis have been extended to cover actual physical structures, such as those visible in various forms of imaging. By analyzing anatomical structures' connectivity and relationships, topology helps identify disease patterns, predict outcomes, and optimize treatment strategies through network-based modeling [27,34]. Thus, in general, *algebraic topology* connects geometric spaces to algebraic structures, while *differential geometry* examines curves and surfaces via differential calculus [27,34,35]. These methods are similar in that they analyze spaces and structures, yet they differ in methods and applications. Unique areas include topological data analysis and its flagship algorithm, persistent homology, topological signatures in algebraic topology, and fractal geometry in differential geometry.

Topological data analysis (TDA) is an umbrella term for a set of methods that aim to make topological information in data sets apparent [36-40]. The flagship method of TDA is termed, "persistent homology," a method that provides an intuitive view on topological features in the data at multiple scales, described by a scale parameter $\in$ (epsilon). The scale parameter

defines the radii of balls expanding from the data points (show in Figure 1). Betti numbers enumerate the features that exist at different dimension levels (connected components, cycles, voids, higher dimensional voids).

For ∈=0 in Figure 1, all points in the dataset are considered disconnected. As ∈ grows, data points become progressively connected to each other, giving birth to features at the 0^{th} Betti number level. Importantly, while the diagram shows this process at the 0^{th} Betti number, any Betti number can change as ∈ grows. Persistent homology tracks the "evolution" of shapes across Betti numbers as ∈ increases; increasing ∈ gives rise to a filtration across the dataset. The crucial insight of persistent homology is that there is no one scale to consider data, but rather, that one should consider data at all scales, thus tracking the changes in shape that are characteristic across all of them. Fig. 1 illustrates this process by means of a simple circular dataset. With a small ∈ parameter, i.e., from a close distance, the circular structure of the data is not apparent. As we "zoom out," though, increasing the ∈ parameter, the circular structure becomes apparent and data points become connected with each other. Topological features that "survive" to a given parameter value ∈ are termed "persistent." While Figure 1 illustrates how increasing ∈ affects the connectedness of a point cloud, this concept can also be readily applied to medical images when investigating the connectedness of pixel data.

In the example shown here (Fig. 1), a cycle is the feature that is persistent across a range of ∈ parameters, appearing at ∈ value 0.2 and disappearing at 1.0. While not apparent from this simple example, topological features afford different interpretations depending on their dimensionality. In low dimensions, topological features correspond to connected components, cycles, and voids, but their mathematical description generalizes to spaces of arbitrary dimension that elude our intuition. Computationally, this process can be applied to point clouds and only involves matrix operations, for example, Gaussian elimination, a technique for reducing the number of equations and variables required to characterize the data. It is possible, however, to generalize and extend persistent homology to a wide variety of modalities, including images.

Regardless of the modality and the computational details, the multi-scale topological features identified and extracted via persistent homology are typically summarized in topological descriptors such as *persistence diagrams*. A d-dimensional persistence diagram is a set of points in the plane, with each point (a, b) representing the *scale* of a topological feature. Often, a is referred to as the "creation" or "birth" time, while b is called the "destruction" or "death" time of the feature. The absolute difference between those two, i.e., $|b - a|$, is called known as the *persistence* of the topological feature. Features of *high persistence* are typically considered to be more relevant or trustworthy (i.e., likely to be of biological/clinical significance as they recur throughout data) than features of *low persistence*, which are typically seen as topological noise [36,37]. These are described in more detail below.

One limitation of current radiomics and pathomics tools is the ever-present potential for overfitting, whereby models may perform exceptionally well on the training data but fail to

generalize to new patient populations. Fortunately, this issue can be mitigated using techniques based on persistent homology, which, because they allow for the identification of stable features in complex data sets, reduce the risk of overfitting by providing features that generalize better. Furthermore, radiomics and pathomics often rely on predefined image and tissue features; while these improve explainability, especially for AI approaches, they can limit their adaptability to evolving medical knowledge. Topology and geometry offer the means to develop dynamic and data-driven approaches that can respond to emerging trends in healthcare.

Persistent Homology – Applications in Radiomics and Pathomics

Persistent homology is a powerful method that has demonstrated utility in the analysis of medical images and histopathology slides [41-43]. Persistent *homology* is concerned with the study of topological features that persist across multiple spatial scales in a dataset. Persistence *diagrams* summarize the birth and death points within a filtration for topological features in the data. Persistence diagrams allow for easy visualization of topological features and can be compared through straightforward distance metrics, such as the Wasserstein distance, to discern statistically different images based on persistent features.

Persistence Landscapes

The persistence *landscapes* approach builds upon this theory by generating persistence diagrams that capture the "birth" and "death" of topological features; these are viewed as a series of landscapes. These landscapes provide a comprehensive vector-based representation of the topological structure of the data [41,42], which permits integration of persistent feature summaries with many common machine-learning methods. In addition to implicit representations, so-called *kernels* [44], there are other efficient methods for obtaining vectorial representations of persistence diagrams. Three common approaches, visualized in Fig. 2, are *persistence images* [45], *Betti curves* [45,46], and *persistence landscapes* [47]. Except for the persistence landscapes, all transformations are inherently *lossy*, i.e., it is not possible in general to reconstruct the original persistence diagram from a persistence image or from a Betti curve. Nevertheless, all these descriptors can be quickly calculated and are well-suited to integration into standard data science tools. They can thus be used to classify data based on topological features, find anomalies, predict outcomes, and more.

The persistence landscape approach has been applied to extract topological features from medical images [47,48]. When assessing the topology of glioblastoma, for example, *homology* would explain the interconnectedness of pixel data in 2D or 3D volumes. Alternatively, *persistent homology* would provide information on the interconnectedness of the pixel data in 2D or 3D volume across varying scale parameters, assessed, for example, by increasing the radius surrounding segmented data points. Finally, *persistence landscapes* can help visually explain the multi-scale topological features obtained from persistent homology by depicting how interconnectedness changes at different scales. Such persistence landscapes can be readily integrated with classical machine learning or deep learning algorithms.

Persistence Landscapes – Applications in Medicine

For example, in the analysis of breast cancer lesions, persistence landscapes have been employed to extract radiomics features that characterize the spatial distribution of tumor subregions and their connectivity, metrics usable for predicting treatment response and patient outcomes [49]. Similarly, in lung cancer, persistence landscapes have been used to quantify the spatial relationships between tumor regions, allowing for the identification of high-risk subregions [6] and personalized treatment strategies. Gao et al. introduced an innovative algorithm to segment high-resolution CT images of cardiac left ventricles, focusing on complex papillary muscle and trabeculae features [50]. By utilizing methods from computational topology, including persistent homology, their algorithm identified missing topological structures, improving segmentation performance. Wu et al. addressed the challenge of reconstructing ventricular trabeculae in cardiac image analysis [51]. Their novel approach detected salient topological handles, refined by a classifier, enhancing segmentation compared to traditional methods, emphasizing the value of topological priors in cardiac image analysis.

In pathomics, persistence landscapes have been used to analyze digitized histopathology slides, converting morphological features into persistence landscapes to provide quantitative representations of tissue morphology and cellular structures. This allows for the identification of biomarkers and patterns that are associated with disease progression and patient outcomes [9]. Persistent homology has also been applied to prostate cancer histopathology, clustering architectural subtypes independently of Gleason patterns. These topological representations offer higher granularity and reproducibility, making them valuable inputs for machine learning methods aimed at enhancing prostate cancer diagnosis and prognosis [6,52]. The derived pathomics features have also shown potential in predicting tumor aggressiveness and guiding treatment decisions. In addition, the persistence landscape approach has been used to analyze immunohistochemistry slides, allowing for the quantification of biomarker expression patterns and the identification of subtypes with different molecular characteristics [53,54]. One recent study using pathology images has demonstrated utility in combining persistent homology profiles and CNNs to classify specimens from tumor-bearing and normal patients [55].

Topological approaches, specifically topological image modification (TIM) and topological image processing (TPI), have been shown to enhance object detection and characterization of skin lesions in clinical images (Fig. 4), offering an efficient, unsupervised approach to isolate significant objects within relevant regions [56]. Topological approaches can also track spatial data of cell locations over time to predict malignant behaviors in tumors, offering insights into the intricate patterns and dynamics between tumor and immune cells, and accurately identifying early signs of perivascular niche formation, a proxy for metastasis [57]. The integration of AI with advanced analytics of histological images can also provide insights into the tumor microenvironment (TME). Various deep-learning algorithms including attention-based and multimodal models have been applied for characterizing TME patterns, linking image features with clinical outcomes. Although the majority of AI models are evaluated retrospectively, newly available datasets and increasing computational power can

assist in prospective validation and demonstration of actual benefits in both statistical and clinical modes [58].

Abousamra et al. explored digital pathology's intricacies by utilizing spatial statistics and topological data analysis to model cell contexts [59]. They introduced new mathematical tools combining a location-specific function with topological features, enabling the generation of high-quality multi-class cell layouts. These layouts, a novelty in the field, demonstrated potential for data augmentation and improving cell classification. Lawson et al. pioneered prostate cancer histology assessment with topological data analysis, clustering images into architectural groups beyond familiar Gleason patterns [52]. Their persistent homology approach showed sensitivity in identifying sub-architectural groups, offering a refined quantification method with applications in diagnosis and prognosis determination.

Topological Signatures

Topological signatures involving homology profiles represent another promising method that leverages concepts from algebraic topology to characterize topological properties of complex data, providing valuable insights into disease characterization and prediction. As touched on above, rather than converting a persistence diagram into a vector-compatible format to integrate with other datasets, as persistence landscapes do, topological signatures capture a vector of features in an image, focusing less on start and end appearances. These signatures offer a compact summary of the "topological activity" within the data filtration, improving image analysis, for instance [60]. While persistence landscapes map how the interconnectedness of data points change over differing scales, measuring when they appear and disappear, topological signatures typically focus on the total number of new connections that can be formed. Long-lived topological features are captured, stored, and then parsed into a relational database as either a vector feature or as individual features, with a number given for each column representing a feature dimension.

Despite its promise, however, persistence landscape-based approaches are challenged by issues involving computational hurdles and complex interpretation. Additionally, the selection of appropriate parameters and the standardization of persistence landscapes across various datasets and applications require care.

Topological Signatures – Applications in Medicine

These signatures can help determine more comprehensive and robust predictive tumor characterization and patient stratification [27]. In the case of gliomas, topological features can distinguish between different subtypes as well as predict patient survival [34]. Similarly, for breast cancer, the approach has been utilized to analyze mammographic images and extract topological features related to the spatial arrangement and connectivity of microcalcifications, aiding in the detection and characterization of malignancies [61,62]. Fig. 5 illustrates three groups of brain scans: a young (top left), aging (middle left), and Alzheimer's disease patient (bottom left), along with a diagram of filtration features for each group. Note the smaller peak found in the young group.

In prostate cancer grading, a histology-centric task, it is possible to extract topological features that capture architectural patterns and glandular structures that encapsulate differences between low- and high-grade tumors. A 2020 study by Yan et al. introduced statistical representations of homology profiles (SRHP), a persistent-homology-based method, for automating Gleason grading in prostate cancer assessment [63]. They achieved a remarkable accuracy of 0.89 and an AUC of 0.96 in distinguishing between Grade 3 and Grade 4 patterns on prostate biopsy slides. What sets SRHP apart is its ability to provide not only effective results but also interpretability, aligning well with the practices of pathologists (Fig. 6). This advancement holds significant promise for enhancing the accuracy and consistency of prostate cancer diagnosis and prognosis.

In lymph node analysis, a topological signatures approach has been used to extract features that correlate with metastatic potential, aiding in the identification of high-risk patients [64]. However, the vectorization of topological features in persistence diagrams requires significant computational resources and application of robust algorithms. The interpretation and integration of topological features in conjunction with other data modalities, such as genomics and clinical data, is also non-trivial. However, by integrating mathematical tools from spatial statistics and topological data analysis into a deep generative model, it becomes possible to create high-quality multi-class cell layouts. These topology-rich layouts enhance data augmentation, significantly improving downstream tasks such as cell classification [59].

Geometric Feature Extraction – Applications in Radiomics and Pathomics

While tools from topology capture global properties that are not affected by stretching, squishing, or other image manipulations, geometric properties such as curvature, distance between points, and volumes of objects change under such distortions of the underlying data manifold. However, in many cases, geometric features are important to capture. For instance, thickening of a blood vessel or development of plaques in that blood vessel do not change the *topology* of the vessel: it is still a tube with a hollow interior (unless the vessel is fully occluded to eliminate that hollow tube). Yet, the *geometry* certainly changes and may be of interest when studying the progression of stenosis and atherosclerosis. An aneurysm's bubble within a vessel, similarly, does not change the fact that the vessel is still a tube. However, the bulge does affect vessel geometry and impacts blood flow and patient risk. Given these differences, many geometry-based feature extraction tools exist, and geometry-based image modelling is a growing field in medical imaging.

The geometric feature extraction (GFE) approach is a powerful method used in radiomics and pathomics for the analysis of medical images and histopathology slides. This approach focuses on extracting quantitative geometric features that capture the spatial characteristics and shape properties of anatomical structures and cellular components [65,66], taking advantage of the wide range of quantitative features it generates. These features can be computed from region-based or contour-based segmentation approaches, depending on the specific task and dataset. Various geometric descriptors, including size-based features (e.g., volume, area, diameter), shape-based features (e.g., circularity, elongation, complexity), and spatial

distribution-based features (e.g., distance, clustering), can be extracted to capture various aspects of anatomical or cellular structures [27].

Fusion of deep convolutional (CNN) and geometric features is an approach that increases boundary contrasts and enhances features, which can then be deployed downstream in CNN-based modelling. This approach, applied to capsule-based endoscopy images, was shown to decrease training time five-fold compared to existing state-of-the-art methods, without sacrificing performance quality [67]. Recent advancements in object recognition, such as hyperbolic visual embedding, which maps flat images onto hyperbolic spaces prior to modelling, have also shown promise in distinguishing features that exist in an image or set of images. While this has not yet been applied to pathomics, the potential for recognizing many types of cells and cellular aberrations with this method is promising [67]. In all, differential geometry distinguishes itself from standard spatial analysis tools by seamlessly integrating basic methods, such as size-based and shape-based feature extraction, with advanced techniques such as hyperbolic or fractal mapping. This convergence not only optimizes compute time and accuracy but also paves the way for medical models to scale efficiently and incorporate diverse data sources [68].

Integration with Other Data Sources

Features arising just from imaging studies or biometric measurements often do not capture all relevant information that could contribute to diagnosis or modelling disease progression. Multimodal models including patient histories, biometric measurements, image features and other electronic health record data constitute a more complete biopsychosocial view of disease and patient histories. Topological signatures and persistence landscapes integrate well with other types of data—such as genomics data, biometric records, and extractions from clinical notes, as these can be expressed as vectors. Often, such combined models provide better insights and better patient management opportunities than models that only include one type of data (such as imaging) [69]. However, integration of text features, imaging features, and relational data from electronic health record systems (such as biometric records or medication lists) is complicated. Thus, tools such as topological signatures that allow medical images to integrate easily with other data types will be helpful in creating powerful multimodal models at scale. However, caution is still required when multi-dimensional data is used to generate biological or clinical conclusions, because the opportunity of overfitting increases as the number of variables increases, potentially yielding results that may be misleading or entirely spurious. Though it is a prevalent belief that cross-validation acts as a safeguard against overfitting or "overhyping," this is not always the case. Random data can still produce spurious results even when cross-validation is employed, for example by merely tweaking hyperparameters in ways that might otherwise seem harmless [70].

Genomics

Genomics approaches consist of data that can be processed by many machine learning methods to identify rare mutations, common tumor profiles, and other relevant diagnostic and prognostic features for modelling. Data topology approaches have played a role recently in tracking changes in tumor genetics over time [71] and in identifying genetic alterations and

markers in tumors [72]. Certainly, one of the fastest growing fields today is radiogenomics, in which tumor imaging profiles are linked to tumor genetic profiles [73]. By linking imaging with specific genetic profiles of tumors, it is possible to discern new cancer phenotypes and develop specialized treatments for rare tumor subtypes. However, the lack of accessible imaging and accessible genomics data often hinders collaboration between radiologists, pathologists, and geneticists studying the same diseases or working on management of the same patients [74]. For example, the use of radiogenomics has led to successful prediction of glioblastoma survival time by modelling patient imaging data (via geometric feature extraction) and tumor genetic profiles; the combination can outperform survival models that use only imaging data or only tumor genetic profiles [75]. It is possible that persistence landscapes and topological signatures can also play a role in integrating such data with the pathomics and radiomics information outlined above.

Future Perspectives

The potential of radiomics and pathomics tools to contribute to the goals of precision medicine is evident. As these fields continue to advance, we can expect exciting developments that will shape the landscape of medical imaging and histopathology analysis. The integration of radiomics and pathomics with other -omics data, such as genomics, proteomics, and metabolomics, along with other patient-associated data, will enrich our understanding and the utility of these approaches, leading to more accurate diagnosis, personalized treatment selection, and improved outcomes. Machine learning and artificial intelligence algorithms, particularly those rooted in geometry and topology, can play a crucial role in this process. The development of standardized protocols and guidelines for radiomics and pathomics analysis can address some reproducibility and generalizability issues, enabling increasingly diverse datasets to be explored and facilitating multicenter collaborations. Finally, a potential advantage is that geometry and/or topology provide additional prior information, which may obviate the need to collect ever-larger data sets. Together, these tools may help define the shape of things to come.

References:

1. Jayatilake S, Ganegoda GU. Involvement of Machine Learning Tools in Healthcare Decision Making. *J Healthc Eng*. 2021;2021:6679512.
2. Johnson KB, Wei WQ, Weeraratne D, et al. Precision Medicine, AI, and the Future of Personalized Health Care. *Clin Transl Sci*. Jan 2021;14(1):86-93.
3. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. Feb 2016;278(2):563-77.
4. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. *Nat Rev Cancer*. Aug 2018;18(8):500-510.
5. van Timmeren JE, Cester D, Tanadini-Lang S, et al. Radiomics in medical imaging-"how-to" guide and critical reflection. *Insights Imaging*. Aug 12 2020;11(1):91.
6. Mashatola L, Akhalwaya I, Muller S. Topological Data Analysis-Deep Learning Framework for Predicting Cancer Phenotypes. Open Review: IBM Research Africa; 2022.
7. Brancato V, Cavaliere C, Garbino N. et al. The relationship between radiomics and pathomics in Glioblastoma patients: Preliminary results from a cross-scale association study. *Front Oncol*. 2022;12:1005805.
8. Gupta R, Kurc T, Sharma A. et al. The Emergence of Pathomics. *Current Pathobiology Reports*. 2019;7:73–84.
9. Lu C, Shiradkar R, Liu Z. Integrating pathomics with radiomics and genomics for cancer prognosis: A brief review. *Chin J Cancer Res*. Oct 31 2021;33(5):563-573.
10. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. Dec 2017;14(12):749-762.
11. Kiehl TR. Digital and Computational Pathology: A Specialty Reimagined. In: Ehsani S, Glauner P, Plugmann P, Thieringer FM, eds. *The Future Cycle of Healthcare*. Springer, Cham; 2022:227–250.
12. Sharma A, Tarbox L, Kurc T, et al. PRISM: A Platform for Imaging in Precision Medicine. *JCO Clin Cancer Inform*. Jun 2020;4:491-499.
13. Wu Y, Cheng M, Huang S, et al. Recent Advances of Deep Learning for Computational Histopathology: Principles and Applications. *Cancers (Basel)*. Feb 25 2022;14(5)
14. Sarker IH. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Comput Sci*. 2021;2(6):420.
15. Jorgensen JT. Twenty Years with Personalized Medicine: Past, Present, and Future of Individualized Pharmacotherapy. *Oncologist*. Jul 2019;24(7):e432-e440.
16. Seyhan AA, Carini C. Are innovation and new technologies in precision medicine paving a new era in patients centric care? *J Transl Med*. Apr 5 2019;17(1):114.
17. Strianese O, Rizzo F, Ciccarelli M, et al. Precision and Personalized Medicine: How Genomic Approach Improves the Management of Cardiovascular and Neurodegenerative Disease. *Genes (Basel)*. Jul 6 2020;11(7)
18. Brittain HK, Scott R, Thomas E. The rise of the genome and personalised medicine. *Clin Med (Lond)*. Dec 2017;17(6):545-551.
19. De Maria Marchiano R, Di Sante G, Piro G, et al. Translational Research in the Era of Precision Medicine: Where We Are and Where We Will Go. *J Pers Med*. Mar 18 2021;11(3)
20. Hartl D, de Luca V, Kostikova A, et al. Translational precision medicine: an industry perspective. *J Transl Med*. Jun 5 2021;19(1):245.
21. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging*. Nov 2012;30(9):1323-41.

22. Huang W, Randhawa R, Jain P, et al. Development and Validation of an Artificial Intelligence-Powered Platform for Prostate Cancer Grading and Quantification. *JAMA Netw Open*. Nov 1 2021;4(11):e2132554.
23. Atkinson C, Ganeshan B, Endozo R, et al. Radiomics-Based Texture Analysis of (68)Ga-DOTATATE Positron Emission Tomography and Computed Tomography Images as a Prognostic Biomarker in Adults With Neuroendocrine Cancers Treated With (177)Lu-DOTATATE. *Front Oncol*. 2021;11:686235.
24. Chiesa-Estomba CM, Echaniz O, Larruscain E. et al. Radiomics and Texture Analysis in Laryngeal Cancer. Looking for New Frontiers in Precision Medicine through Imaging Analysis. *Cancers (Basel)*. Sep 20 2019;11(10)
25. Gao B, Dong D, Zhang H, et al. Editorial: Radiomics Advances Precision Medicine. *Front Oncol*. 2022;12:853948.
26. Liu Z, Wang S, Dong D, et al. The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. *Theranostics*. 2019;9(5):1303-1322.
27. Singh Y, Farrelly CM, Hathaway QA, et al. Topological data analysis in medical imaging: current state of the art. *Insights Imaging*. Apr 1 2023;14(1):58.
28. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med*. Oct 2018;24(10):1559-1567.
29. Wang S, Yang DM, Rong R, et al. Artificial Intelligence in Lung Cancer Pathology Image Analysis. *Cancers (Basel)*. Oct 28 2019;11(11)
30. Wei JW, Tafe LJ, Linnik YA. et al. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Sci Rep*. Mar 4 2019;9(1):3358.
31. Yu KH, Wang F, Berry GJ, et al. Classifying non-small cell lung cancer types and transcriptomic subtypes using convolutional neural networks. *J Am Med Inform Assoc*. May 1 2020;27(5):757-769.
32. Levy-Jurgenson A, Tekpli X, Kristensen VN. et al. Spatial transcriptomics inferred from pathology whole-slide images links tumor heterogeneity to survival in breast and lung cancer. *Sci Rep*. Nov 2 2020;10(1):18802.
33. Van Herck Y, Antoranz A, Andhari MD, et al. Multiplexed Immunohistochemistry and Digital Pathology as the Foundation for Next-Generation Pathology in Melanoma: Methodological Comparison and Future Clinical Applications. *Front Oncol*. 2021;11:636681.
34. Paul R. Topological features in addition to radiomics signature predict 1p19q status and tumor grade in low-grade gliomas. *Imaging System and Technology*. 2021;32(3):753-766.
35. Singh Y, Farrelly C, Hathaway QA, et al. The Role of Geometry in Convolutional Neural Networks for Medical Imaging. *Mayo Clinic Proceedings: Digital Health*. 2023;1(4):519-526.
36. Edelsbrunner H, Harer J. *Computational Topology: An Introduction*. American Mathematical Society; 2009.
37. Edelsbrunner, Letscher, Zomorodian. Topological Persistence and Simplification. *Discrete & Computational Geometry*. 2002;28:511–533.
38. Carlsson G. Topology and Data. *Bulletin of the American Mathematical Society*. 2009;46(2):255–308.
39. Chazal F, Michel B. An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. *Front Artif Intell*. 2021;4:667963.
40. Hensel F, Moor M, Rieck B. A Survey of Topological Machine Learning Methods. *Front Artif Intell*. 2021;4:681108.

41. Fugacci U, Scaramuccia S, Iuricich F. et al. *Persistent homology: a step-by-step introduction for newcomers*. Smart Tools and Apps for Graphics - Eurographics Italian Chapter Conference. The Eurographics Association; 2016.
42. Rammal A, Assaf R, Goupil A. et al. Machine learning techniques on homological persistence features for prostate cancer diagnosis. *BMC Bioinformatics*. Nov 12 2022;23(1):476.
43. Chen S. *Topological Data Analysis for Medical Imaging and RNA Data Analysis on Tree Spaces*. Florida State University; 2021. https://purl.lib.fsu.edu/diginole/2020_Summer_Fall_Shen_fsu_0071E_16291
44. Reininghaus J, Huber S, Bauer U. et al. A Stable Multi-Scale Kernel for Topological Machine Learning. presented at: IEEE Conference on Computer Vision and Pattern Recognition 2015; Red Hook, NY, USA.
45. Adams H, Emerson T, Kirby M, et al. Persistence Images: A Stable Vector Representation of Persistent Homology. *Journal of Machine Learning Research*. 2017;18(8):1−35.
46. Bastian R, Sadlo F, Leitte H. Topological Machine Learning with Persistence Indicator Functions. presented at: Topological Methods in Data Analysis and Visualization; 2020; Cham, Switzerland:.
47. Bubenik P. Statistical Topological Data Analysis using Persistence Landscapes. *Journal of Machine Learning Research*. 2015;16:77-102.
48. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. Mar 2012;48(4):441-6.
49. Asaad A, Ali D, Majeed T. et al. Persistent Homology for Breast Tumor Classification Using Mammogram Scans. *Mathematics*. 2022;10(21)
50. Gao M, Chen C, Zhang S. et al. Segmenting the papillary muscles and the trabeculae from high resolution cardiac CT through restoration of topological handles. *Inf Process Med Imaging*. 2013;23:184-95.
51. Wu P, Chen C, Wang Y, et al. Optimal topological cycles and their application in cardiac trabeculae restoration. presented at: Information Processing in Medical Imaging 2017;
52. Lawson P, Sholl AB, Brown JQ. et al. Persistent Homology for the Quantitative Evaluation of Architectural Features in Prostate Cancer Histology. *Sci Rep*. Feb 4 2019;9(1):1139.
53. Harmon SA, Tuncer S, Sanford T. et al. Artificial intelligence at the intersection of pathology and radiology in prostate cancer. *Diagn Interv Radiol*. May 2019;25(3):183-188.
54. Li H, Lee CH, Chia D. et al. Machine Learning in Prostate MRI for Prostate Cancer: Current Status and Future Opportunities. *Diagnostics (Basel)*. Jan 24 2022;12(2)
55. Qaiser T, Sirinukunwattana K, Nakane K. et al. Persistent Homology for Fast Tumor Segmentation in Whole Slide Histology Images. *Procedia Computer Science*. 2016;90:119-124.
56. Vandaele R, Nervo GA, Gevaert O. Topological image modification for object detection and topological image processing of skin lesions. *Sci Rep*. Dec 3 2020;10(1):21061.
57. Yang J, Fang H, Dhesi J, et al. Topological classification of tumour-immune interactions and dynamics. Cell Behavior. arXiv2023.
58. Fu X, Sahai E, Wilkins A. Application of digital pathology-based advanced analytics of tumour microenvironment organisation to predict prognosis and therapeutic response. *J Pathol*. Aug 2023;260(5):578-591.

59. Abousamra S, Gupta R, Kurc T. et al. Topology-Guided Multi-Class Cell Context Generation for Digital Pathology. presented at: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2023; Vancouver, BC, Canada.
60. Saadat-Yazdi A, Andreeva R, Sarkar R. Topological Detection of Alzheimer's Disease Using Betti Curves. presented at: Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data; 2021;
61. Wang F, Kapse S, Liu S. et al. TopoTxR: A Topological Biomarker for Predicting Treatment Response in Breast Cancer. presented at: Information Processing in Medical Imaging; 2021;
62. de Brito Silva TF, de Paiva AC, Silva AC. et al. Classification of breast masses in mammograms using geometric and topological feature maps and shape distribution. *Research on Biomedical Engineering*. 2020;36:225–235.
63. Yan C, Nakane K, Wang X, et al. Automated gleason grading on prostate biopsy slides by statistical representations of homology profile. *Comput Methods Programs Biomed*. Oct 2020;194:105528.
64. Failmezger H, Muralidhar S, Rullan A. et al. Topological Tumor Graphs: A Graph-Based Spatial Model to Infer Stromal Recruitment for Immunosuppression in Melanoma Histology. *Cancer Res*. Mar 1 2020;80(5):1199-1209.
65. Zhang X, Zhang Y, Zhang G, et al. Deep Learning With Radiomics for Disease Diagnosis and Treatment: Challenges and Potential. *Front Oncol*. 2022;12:773840.
66. Barisoni L, Lafata KJ, Hewitt SM. et al. Digital pathology and computational image analysis in nephropathology. *Nat Rev Nephrol*. Nov 2020;16(11):669-685.
67. Sharif M, Khan MA, Rashid M. et al. Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images. *Journal of Experimental & Theoretical Artificial Intelligence*. 2021;33(4):577-599.
68. Molaei S, Ghorbani N, Dashtiahangar F. et al. FDCNet: Presentation of the Fuzzy CNN and Fractal Feature Extraction for Detection and Classification of Tumors. *Comput Intell Neurosci*. 2022;2022:7543429.
69. Acosta JN, Falcone GJ, Rajpurkar P. et al. Multimodal biomedical AI. *Nat Med*. Sep 2022;28(9):1773-1784.
70. Hosseini M, Powell M, Collins J, et al. I tried a bunch of things: The dangers of unexpected overfitting in classification of brain data. *Neurosci Biobehav Rev*. Dec 2020;119:456-467.
71. Bukkuri A, Andor N, Darcy IK. Applications of Topological Data Analysis in Oncology. *Front Artif Intell*. 2021;4:659037.
72. Rabadan R, Mohamedi Y, Rubin U, et al. Identification of relevant genetic alterations in cancer using topological data analysis. *Nat Commun*. Jul 30 2020;11(1):3808.
73. Bodalal Z, Trebeschi S, Nguyen-Kim TDL, Schats W, Beets-Tan R. Radiogenomics: bridging imaging and genomics. *Abdom Radiol (NY)*. Jun 2019;44(6):1960-1984.
74. Lundstrom CF, Gilmore HL, Ros PR. Integrated Diagnostics: The Computational Revolution Catalyzing Cross-disciplinary Practices in Radiology, Pathology, and Genomics. *Radiology*. Oct 2017;285(1):12-15.
75. Wijethilake N, Islam M, Ren H. Radiogenomics model for overall survival prediction of glioblastoma. *Med Biol Eng Comput*. Aug 2020;58(8):1767-1777.

ACKNOWLEDGEMENTS
We thank our colleagues for support and useful insights.

AUTHOR CONTRIBUTIONS
RML, YS, QAH, and CF were primarily responsible for manuscript content. BR, JR, and GC provided technical material support. PP, AC, DD and YS reviewed the quality of the manuscript. All authors read and approved the final manuscript.

COMPETING INTERESTS
The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE
Not applicable.

Figure Legends

Fig. 1: An overview of TDA and persistent homology, with $\in$ indicating the local scale parameter, and β_1 indicating the number of cycles "detected" at the given scale.

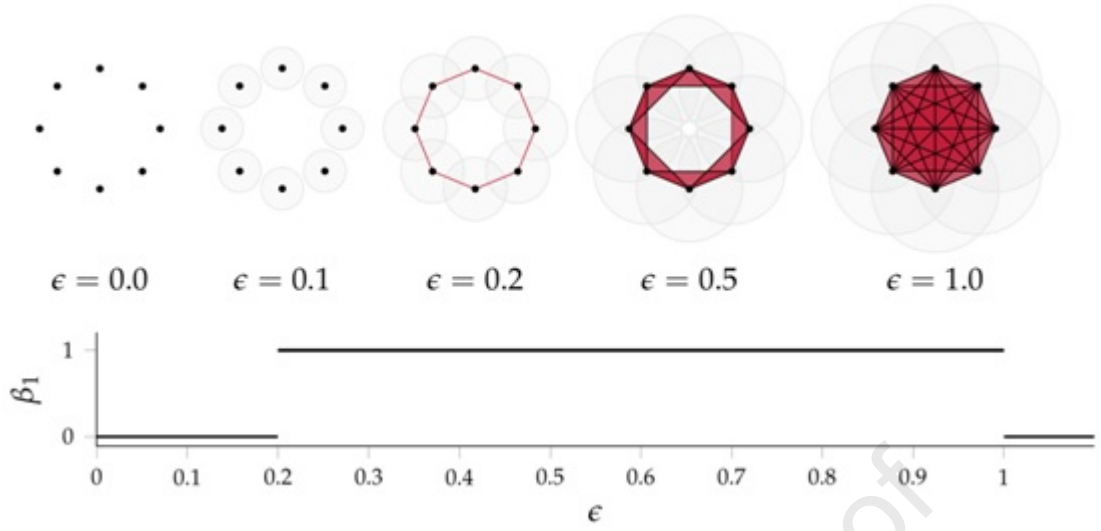
Fig. 2: An overview of different representations of persistence diagrams[40]
(a) Persistence images (on right) are constructed by "smoothing" the persistence diagrams (scatter plots on left). They provide a fast and effective way of using topological features for subsequent data analysis tasks.
(b) Betti curves are a coarse summary of a diagram, obtained by counting the number of active topological features as a function of the overall scale.
(c) Persistence landscapes enjoy stability properties akin to those of persistence diagrams; it is also possible to reconstruct the original features of data from them.
There are no axes for the Persistence images because they are images and only have pixel coordinates.

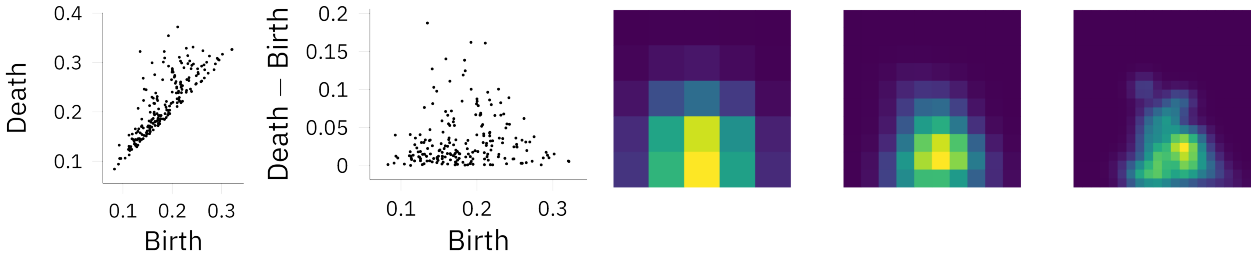
Fig. 3: Algorithm for persistent homology profiles for standard hematoxylin & eosin-stained slides [55].

Fig. 4: Algorithm for persistent homology profiles for gross lesion images[56].

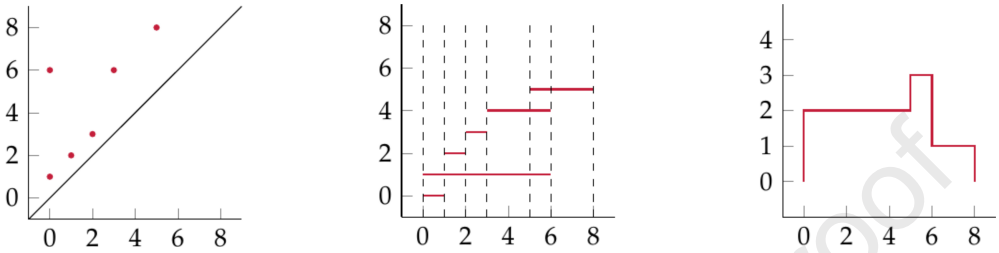
Fig. 5: Topological changes with aging and Alzheimer's disease show the effects in grey matter (red) and white matter (white). Both aging and Alzheimer's disease result in the loss of white matter volume and a thinning of the grey matter. Segmenting colors measures only area, missing intricate topologies. While the Betti number as such only captures coarse summary statistics about data, i.e., the number of connected components or cycles, they can be evaluated alongside a filtration (a growth process) to improve their expressivity. In the form of such *Betti* curves, the Betti numbers capture complex structures, offering deeper, more comprehensive insights than mere area measurements. Grey matter of the aging brain breaks down into smaller connected components, illustrated by the example of a dimension 1 Betti curve computed from the grey matter. Mean subject ages were 56 for the "young" cohort, 83 for the aging cohort, and 65 for patients with Alzheimer's disease.

Fig. 6: This image describes a process involving binary image manipulation and the computation of Betti numbers. Initially, an H-stained component image is transformed into binary format using various threshold values, specifically 40, 100, and 150. The resulting binary images highlight connected components, represented as blue dots (b0), and empty spaces, marked with red stars (b1). This process helps visualize and quantify the structural characteristics of the images in terms of Betti numbers [63].

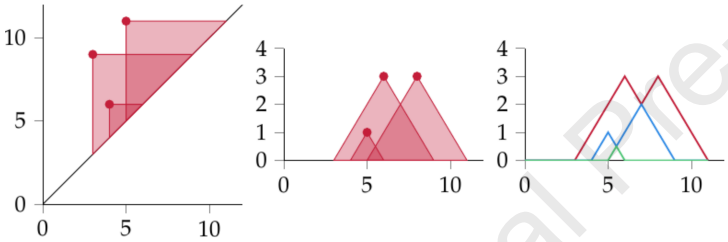




(a) Persistence images construction



(b) Betti curve construction



(c) Persistence landscape construction