

Detecting Temporal shape changes with the Euler Characteristic Transform

LEWIS MARSH

Mathematical Institute & Ludwig Institute for Cancer Research, University of Oxford, Oxford, OX2
6GG, UK

FELIX Y. ZHOU

Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center, Dallas, TX 75235, USA

XIAO QIN

Department of Oncology, University College London, London, WC1E 6DD, UK

XIN LU

Ludwig Institute for Cancer Research, University of Oxford, Oxford, OX3 7DQ, UK

HELEN M. BYRNE

Mathematical Institute & Ludwig Institute for Cancer Research, University of Oxford, Oxford, OX2
6GG, UK

AND

HEATHER A. HARRINGTON*

Mathematical Institute & Wellcome Centre for Human Genetics, University of Oxford, Oxford, OX3
7BN, UK

*Corresponding author: harrington@maths.ox.ac.uk

[Received on 22 December 2022; revised on 18 December 2023; accepted on 16 April 2024]

Organoids are multi-cellular structures that are cultured *in vitro* from stem cells to resemble specific organs (e.g., brain, liver) in their three-dimensional composition. Dynamic changes in the shape and composition of these model systems can be used to understand the effect of mutations and treatments in health and disease. In this paper, we propose a new technique in the field of topological data analysis for DETecting Temporal shape changes with the Euler Characteristic Transform (DETECT). DETECT is a rotationally invariant signature of dynamically changing shapes. We demonstrate our method on a data set of segmented videos of mouse small intestine organoid experiments and show that it outperforms classical shape descriptors. We verify our method on a synthetic organoid data set and illustrate how it generalizes to 3D. We conclude that DETECT offers rigorous quantification of organoids and opens up computationally scalable methods for distinguishing different growth regimes and assessing treatment effects.

Keywords: topological data analysis, oncology, kernel methods, Euler characteristic transform.

1. Introduction

Organoids are three-dimensional multi-cellular *in vitro* cell cultures that can be generated from stem cells. Organoids are also known as *mini organs* because their constituent cells can differentiate into various cell lineages with defined cellular functions. At early times, their growth is approximately radially symmetric. Growth factors in the culture medium surrounding the organoids drive the stem cells to proliferate and the organoids to increase in size. On longer times, the progeny of the stem cells differentiate or specialize into

different cell types that self-organize to produce tissues whose composition resembles the original organ. Excessive cell proliferation combined with cell differentiation can lead to complex shape changes, with the outer boundary of the organoids adopting intricate and asymmetric structures. Organoids have been increasingly used to study tissue development and disease progression. This increase can be attributed to their ability to recapitulate the 3D cellular architecture and function of their tissue of origin than 2D cell cultures (27). The responses of organoids to genetic and environmental perturbations are thereby similar to that of the tissue they were derived from. Organoids are also being used to investigate tissue responses to genetic and environmental perturbations and for drug testing and development (7). While organoids can now be generated to study multiple organs, including the brain, kidney and liver, the most widely studied organoids are those that derive from the intestinal epithelium, one of the fastest-renewing mammalian tissues.

Differentiated cells with specific cellular functions often exhibit defined cellular morphology. Descriptors of morphology can be used as a phenotypic read-out of the organoids, which may reflect the underlying genetic composition and its response to environmental perturbations. Therefore, several studies have focused on quantifying and analysing organoid morphology. Simple measures, such as cell numbers, organoid volume and surface area, diameter, shape factor (ratio of surface area to volume) and growth rate have been used to relate morphology, genotype and drug responses (6; 19; 24; 42). Furthermore, deep learning methods can segment organoid images and extract morphological features including organoid perimeter and eccentricity (18; 23). At the same time, mechanistic models have been developed to investigate the relationship between stem cell proliferation, cell fate specification, organoid growth and morphology. These agent-based and continuum-based models have been compared with experimental data using simple growth curves (19; 37; 42). The complexity of organoid morphology lends itself to more sophisticated analysis. For example, genus and average curvature (20), measures from geometry and topology, have been used to distinguish shape changes in organoids. Here we propose studying the geometry and topology of organoids with topological data analysis (TDA).

Topological data analysis is a collection of data science methods for quantifying shape in data sets by using techniques from topology. TDA has been successfully applied to a variety of problems in applied mathematics, which include sensor network coverage (11), image signalling (4; 8), skeletonizing of images (12), discovering periodicities in time-series (35), material science (25; 26) and financial mathematics (17; 29). A field of applied mathematics in which TDA has been particularly successful is mathematical biology, for example in studying enzyme kinetics (30), neuron morphology (3; 22) and protein structures (13; 15; 31).

One technique in TDA, introduced by Turner, Mukherjee and Boyer (38), is the Euler Characteristic Transform (ECT). The ECT gives a signature for shapes embedded in Euclidean space \mathbb{R}^d . The ECT and its extensions have been applied to a range of problems including MRT scans of glioblastoma patients (9; 21), to classification of 3D shapes (e.g., mammalian teeth (39), protein structures (36) and barley seeds (2)). The above studies highlight the insight that ECT can generate when applied to static data. Here we show the additional information that can be gained from generalizing the ECT to dynamic shape changes (i.e., spatio-temporal data).

1.1 Contributions

We propose an extension to the smooth ECT (SECT), *DEtecting Temporal shape changes with the Euler Characteristic Transform* (DETECT), which is the first application of a rotationally invariant and dynamic ECT. The output from DETECT is a signature in a Hilbert space that can be thought of as a surface in \mathbb{R}^3 . To aid classification of the resulting signature, we apply a non-linear transformation using

kernel methods. In particular, we are the first to use kernel approximation methods in conjunction with the ECT or similar methods. Using such methods, in our case the Nystroem method, makes data analysis pipelines using the ECT or related methods scale to large data sets.

We apply DETECT to dynamic data describing the boundaries of experimental and synthetic organoids in 2D and 3D. We demonstrate the proposed method on the applicability of both synthetic and experimental on 2D and 3D data to distinguish morphology of organoids. First, we show that the smooth Euler Characteristic Transform supersedes standard measures used to study organoid morphology. Next, we show that the proposed temporal method, DETECT, improves the classifying of different organoid treatment groups.

1.2 Related mathematical work

Other studies have successfully used kernels in conjunction with the ECT (9; 36; 39); we are the first to use approximate feature embeddings with the Nystroem method (41). In practice, Nystroem approximations typically scale better in runtime and memory (as $\mathcal{O}(mn)$, where n is the number of data points and $m \ll n$ than kernel methods (41), including Gaussian process approaches with the ECT (9; 36; 39) (which run in $\mathcal{O}(n^3)$). Compared with standard kernel methods, these approximations also allow a wider variety of methods to be used with DETECT, including random forests.

1.3 Structure

Our paper is organized as follows. We first introduce relevant background on topological data analysis and kernel methods. We then introduce a novel signature, DETECT, which quantifies the evolution of a shape over time. By applying DETECT to a data set of 2D organoid boundaries segmented from videos of mouse small intestine organoid experiments, we demonstrate that DETECT can distinguish between treated and untreated organoids. Our results thereby give insights into how cancer treatments affect the morphology of organoids. We highlight that DETECT outperforms classical shape descriptors at this classification task. Finally, we demonstrate on a data set of 3D organoid boundaries generated by a mechanistic model of Yan *et al.* (42) that our method generalizes to 3D boundaries and verify that it can extract biologically meaningful information in a setting in which we have precise control over all biologically meaningful factors.

2. Mathematical methods

2.1 Preliminaries

2.1.1 Simplicial complexes and filtrations. To undertake our work, we require a mathematical definition of a shape. In topological data analysis (TDA) and computational geometry, simplicial complexes are widely used for this purpose:

DEFINITION 2.1 Given a finite set of vertices X , an *abstract simplicial complex* is a set of subsets of X , denoted \mathcal{K} , such that for any $\tau \in \mathcal{K}$ and $\sigma \subseteq \tau$, we have $\sigma \in \mathcal{K}$. We call $\sigma \in \mathcal{K}$ a *simplex* of \mathcal{K} . Moreover, for any simplex σ , we define $\dim(\sigma) = |\sigma| - 1$ and $\mathcal{K}_n = \{\sigma \in \mathcal{K} \mid \dim(\sigma) = n\}$. We call the simplices in \mathcal{K}_n n -simplices.

A *geometric (or embedded) simplicial complex* \mathcal{K} is an abstract simplicial complex that is endowed with an embedding in \mathbb{R}^d . That is, $X \subset \mathbb{R}^d$ and for all $\sigma = \{x_0, \dots, x_k\} \in \mathcal{K}$ we have that x_0, \dots, x_k are affinely independent, meaning that $x_1 - x_0, \dots, x_k - x_0$ are linearly independent. We can then think of \mathcal{K} equivalently as the union of the convex hulls of all of its simplices, which is a topological subspace of \mathbb{R}^d .



FIG. 1. Example of two geometric simplicial complexes, \mathcal{K}_1 and \mathcal{K}_2 , embedded in \mathbb{R}^2 . Vertices as blue dots, 1-simplices as red lines and 2-simplices as grey areas.

Given two simplicial complexes \mathcal{K} and \mathcal{K}' , a *simplicial map* f is a function $f : X \rightarrow X'$, extending to a map $f : \mathcal{K} \rightarrow \mathcal{K}'$ by $f(\sigma) = \{f(x) \mid x \in \sigma\}$.

An example of two embedded simplicial complexes is given in Fig. 1.

Note that any abstract simplicial complex can be viewed as a geometric simplicial complex by considering X to be a subset of the free vector space generated by X . This geometric simplicial complex is called the *geometric realization* of \mathcal{K} .

In most settings, abstract simplicial complexes are more amenable to computations while geometric simplicial complexes contain geometric information. The organoid boundaries extensively studied in the following section are all equivalent when modelled as abstract simplicial complexes. We use filtrations to study and compare geometric simplicial complexes representing organoid boundaries.

DEFINITION 2.2 A *filtration* of a simplicial complex \mathcal{K} is a function $f : \mathcal{K} \rightarrow \mathbb{R}$ such that $f(\sigma) \leq f(\tau)$ for $\sigma \subseteq \tau$. For $j \in \mathbb{R}$, we then define

$$\mathcal{K}^j = \{\sigma \in \mathcal{K} \mid f(\sigma) \leq j\}.$$

All \mathcal{K}^j are simplicial complexes and $\mathcal{K}^j \subseteq \mathcal{K}^i$ for $j \leq i$.

Assume that \mathcal{K} is a geometric simplicial complex in \mathbb{R}^d . Let $v \in S^{d-1} = \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ be a fixed direction in \mathbb{R}^d . We then call the filtration on \mathcal{K} induced by

$$f_v : \mathcal{K} \rightarrow \mathbb{R}, \quad \sigma \mapsto \min_{x \in \sigma} \{\langle x, v \rangle\}$$

the *sub-level set filtration* of \mathcal{K} in direction v , where $\langle \cdot, \cdot \rangle$ is the standard inner-product in \mathbb{R}^d . We denote the above filtration $\mathcal{K}^{(\cdot, v)}$ and each sub-level set at $t \in \mathbb{R}$ as $\mathcal{K}^{(\cdot, v) \leq t}$.

We emphasize that not all filtrations of a simplicial complex arise as a sub-level-set filtration.

2.1.2 Kernels and kernel approximations. Before introducing further topology, we will briefly discuss kernels and their associated Hilbert spaces—a theory we will use when defining key topological signatures. Kernels are generalizations of inner-products. The motivation for generalizing inner-products is twofold. First, we may want to use data science methods requiring an inner-product (e.g., support vector classification (SVC), principal component analysis (PCA) or k-means) on data in spaces not endowed with an inner-product. Second, even if it is possible to define an inner-product in data space, features in the data may not be linear. For example, not every labelled data set can be separated by a plane, resulting in inaccurate classification. In such an instance, SVC in combination with kernels could, by contrast, allow a separation.

DEFINITION 2.3 Let \mathcal{X} be a nonempty set. A function $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is called a *kernel* if it is symmetric and positive definite. That is, for any $x_1, \dots, x_n \in \mathcal{X}$ and $a_1, \dots, a_n \in \mathbb{R}$, we get

$$\sum_{i,j=1}^n a_i a_j k(x_i, x_j) \geq 0. \quad (2.1)$$

For general \mathcal{X} , the Kronecker delta gives a kernel. If \mathcal{X} is an inner-product space, the inner-product is a kernel. If \mathcal{X} is a metric space with metric d that is flat, then the function

$$k(x, y) := \exp\left(-\frac{d(x, y)^2}{\lambda}\right),$$

where $\lambda > 0$ is a hyperparameter, is a kernel called the *Gaussian kernel* (14). The Gaussian kernel is the main kernel that we will employ in this paper.

DEFINITION 2.4 Let k be a kernel on some set \mathcal{X} . If we define $\mathcal{H}_0 = \text{span}_{\mathbb{R}}\{k(x, \cdot) \mid x \in \mathcal{X}\}$, then \mathcal{H}_0 is a vector space of functions from \mathcal{X} to \mathbb{R} . Note that

$$\langle k(x, \cdot), k(y, \cdot) \rangle_{\mathcal{H}_0} := k(x, y)$$

defines an inner-product on \mathcal{H}_0 by bi-linear extension. We define $\mathcal{H} = \overline{\mathcal{H}_0}$, the completion of \mathcal{H}_0 . Then \mathcal{H} is a Hilbert space, called the *Reproducing Kernel Hilbert Space* (RKHS) of k .

The name RKHS derives from the fact that for any $f \in \mathcal{H}$ and $x \in \mathcal{X}$, we have $\langle k(x, \cdot), f \rangle_{\mathcal{H}} = f(x)$, which is called the reproducing property of k . In particular, the reproducing property of an RKHS together with the Cauchy–Schwarz inequality gives that the linear functionals $\delta_x : \mathcal{H} \rightarrow \mathbb{R}$ defined by $\delta_x(f) = f(x)$ at $f \in \mathcal{H}$ are continuous for all $x \in \mathcal{X}$.

The main point—in the context of this work—of defining an RKHS is to illustrate that applying a kernel to elements of \mathcal{X} can be viewed as first embedding \mathcal{X} into some Hilbert space of functions \mathcal{H} (by $x \mapsto k(x, \cdot)$) and then taking an inner-product of such embedded elements. While \mathcal{H} may be infinite-dimensional and thus the embedding of \mathcal{X} into \mathcal{H} is intractable in general, we never need to compute the (exact) embedding itself—computing $k(x, y)$ for all $x, y \in \mathcal{X}$ is sufficient for any downstream method relying on the inner-product of \mathcal{H} only. This insight is called the *kernel trick*.

If \mathcal{X} is of finite size n , then computing $k(x, y)$ for all pairs $k(x, y)$ may require only finitely many computations and will scale as $\mathcal{O}(n^2)$. To enable computations for large n , a number of approximation methods of lower computational complexity have been developed. One such method is the Nystroem approximation.

DEFINITION 2.5 Let $\mathcal{X} = \{x_1, \dots, x_n\}$ be of finite size n and let $m < n$, where m is an integer. Denote by K the $n \times n$ -matrix with i, j -entry $k(x_i, x_j)$. Then K can be written in block-form as

$$K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix},$$

where K_{11} is of dimension $m \times m$, K_{22} is of dimension $(n-m) \times (n-m)$ and K_{12} and K_{21} are of dimension $(n-m) \times m$ and $m \times (n-m)$, respectively. For $C := [K_{11}, K_{12}]^T$, the m -th *Nystroem approximation* of

K is the matrix

$$\tilde{K} = CK_{11}^\dagger C^T, \quad (2.2)$$

where † denotes the Moore–Penrose pseudo-inverse.

If K is of approximate rank m (or less), then $\tilde{K} \approx K$ (28). Typically, the approximate rank of a Gram matrix is much less than n if n is large. Computing \tilde{K} only requires $\mathcal{O}(mn)$ evaluations of k , where m is typically fixed. In practice, we only compute $C(K_{11}^\dagger)^{1/2}$ (to save computer memory), as the standard inner-product of the i -th and j -th rows of $C(K_{11}^\dagger)^{1/2}$ approximately gives $k(x_i, x_j)$. This matrix $C(K_{11}^\dagger)^{1/2}$ can be used as a non-linear transformation and can be further analysed. In particular, we take the row vectors and feed them to a method using inner-products. Computing $C(K_{11}^\dagger)^{1/2}$ has a runtime complexity of $\mathcal{O}(nm^2 + m^3)$. The m -dimensional row vectors are also referred to as the *Nystroem embedding* of \mathcal{X} . There exist sampling heuristics for picking an optimal set of m points from \mathcal{X} (28).

2.1.3 The Euler Characteristic and the Euler Characteristic Transform. The *Euler characteristic* is a topological invariant of simplicial complexes. Any two simplicial complexes that are (homotopy) equivalent as topological spaces have the same Euler characteristic. Conversely, if two simplicial complexes have different Euler characteristics, we can conclude that they are topologically different (i.e., not homotopy equivalent). We can compute the Euler characteristic entirely from the combinatorial information of an abstract simplicial complex.

DEFINITION 2.6 Let \mathcal{K} be a simplicial complex. Then its Euler characteristic is

$$\chi(\mathcal{K}) = \sum_{i=0}^{\infty} (-1)^i \cdot |\mathcal{K}_i|. \quad (2.3)$$

Given a geometric simplicial complex \mathcal{K} in \mathbb{R}^d , using a sequence of Euler characteristics induced by a sub-level set filtration yields additional discriminative information compared to $\chi(\mathcal{K})$. Assume that $a \in \mathbb{R}$ is larger than the diameter of \mathcal{K} in \mathbb{R}^d . Then $f_v^{-1}((-\infty, -a)) = \emptyset$ and $f_v^{-1}((-\infty, a)) = \mathcal{K}$ for all $v \in S^{d-1}$. We can now define the Smooth Euler Characteristic transform and related constructions (38).

DEFINITION 2.7 First, let the *Euler characteristic curve* in a fixed direction v be

$$\text{ECC}_{\mathcal{K}}^v : [-a, a] \rightarrow \mathbb{Z}, \quad t \mapsto \chi(\mathcal{K}^{(v, \cdot)^{\leq t}}).$$

Secondly, this curve is smoothed by defining the *smooth Euler characteristic curve* (SEC) as follows.

$$\text{SEC}_{\mathcal{K}}^v : [-a, a] \rightarrow \mathbb{R}, \quad t \mapsto \int_{-a}^t \left(\text{ECC}_{\mathcal{K}}^v(x) - \overline{\text{ECC}_{\mathcal{K}}^v} \right) dx,$$

where $\overline{\text{ECC}_{\mathcal{K}}^v}$ is the mean of the function $\text{ECC}_{\mathcal{K}}^v$ over the interval $[-a, a]$.

Thirdly, the *ECT of a shape* \mathcal{K} is then the map

$$\text{ECT}(\mathcal{K}) : S^{d-1} \rightarrow CF([-a, a]), \quad v \mapsto \text{ECC}_{\mathcal{K}}^v, \quad (2.4)$$

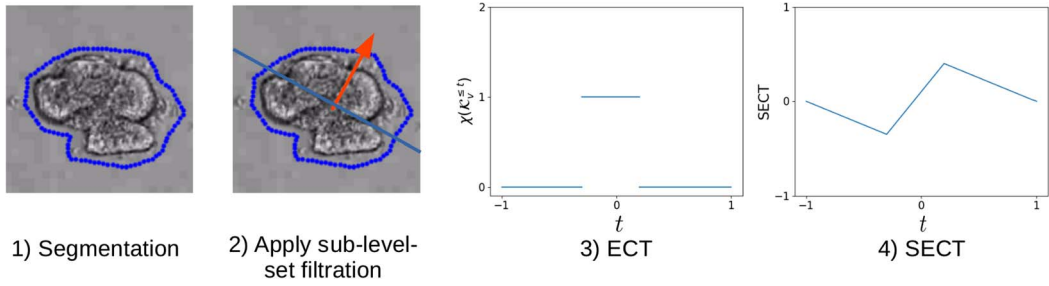


FIG. 2. Standard ECT pipeline visualized on a segmented organoid boundary. 1) Segmented input data given by a geometric simplicial complex \mathcal{K} in blue. 2) Illustration of the sub-level set filtration in the direction v given by the red arrow. The purple line is moved across the blue simplicial complex in the direction of the red arrow. At each offset t of the line, $\mathcal{K}^{(v, \cdot) \leq t}$ is given by the part of \mathcal{K} below the purple line. 3) The ECT in the direction given in 2). 4) The SECT in the direction given in 2).

where $CF([-a, a])$ is the set of constructible integer-valued functions on $[-a, a]$. For a more comprehensive discussion of these constructible functions and ECT, see (10, 5). Finally, we can define the *smooth Euler characteristic transform* (SECT):

$$\text{SECT}(\mathcal{K}) : S^{d-1} \rightarrow L^2([-a, a]), \quad v \mapsto \text{SECT}_{\mathcal{K}}^v. \quad (2.5)$$

Here, $L^2([-a, a])$ is the Hilbert space of all functions on $[-a, a]$ that can be generated by Fourier series.

An overview of the ECT and SECT calculation is given in Fig. 2.

Turner and colleagues have shown that both the ECT and SECT are injective, making them sufficient statistics for comparing shapes embedded in \mathbb{R}^2 and \mathbb{R}^3 (38). Ghrist *et al.* (16) and Curry *et al.* (10) independently extended this injectivity result to general \mathbb{R}^d . In particular, the ECT is able to distinguish distinct geometric simplicial complexes that have the same Euler characteristic. The ECT and SECT are also able to discriminate between shapes that are equivalent up to translation, rotation, reflection and combinations thereof. The issue of discriminating between shapes equivalent up to translation can be overcome by re-centring simplicial complexes by subtracting the mean of all vertices from each vertex in the geometric simplicial complex. However, resolving rotation and reflection requires more care. Fortunately, Curry *et al.* (10) presented a variant of the ECT that is injective on the space of shapes modulo actions of the orthogonal group $O(d)$.

Before introducing their result, we recall the notion of a pushforward measure and define a generic geometric simplicial complex.

DEFINITION 2.8 Let (X_1, Σ_1) and (X_2, Σ_2) be measurable spaces, $f : X_1 \rightarrow X_2$ be a measurable function and $\mu : \Sigma_1 \rightarrow [0, \infty]$ be a measure on X_1 . Then $f_*\mu$, the pushforward of μ along f , is the measure on X_2 defined by $(f_*\mu)(U) = \mu(f^{-1}(U))$ for each $U \in \Sigma_2$.

DEFINITION 2.9 (Definition 6.1 in (10)). Let \mathcal{K} be a geometric simplicial complex in \mathbb{R}^d with vertex set X . We call \mathcal{K} *generic* if

1. the curves $\text{ECT}_{\mathcal{K}}^v$ are distinct for all $v \in S^{d-1}$, and
2. the vertex set X is in general position.

THEOREM 2.10 (Theorem 6.7 in (10)). Let \mathcal{K} and \mathcal{K}' be generic geometric simplicial complexes embedded in \mathbb{R}^d . Let μ be the Lebesgue measure on S^{d-1} . If $\text{ECT}(\mathcal{K})_*(\mu) = \text{ECT}(\mathcal{K}')_*(\mu)$, then there exists a $\phi \in O(d)$ such that $\mathcal{K} = \phi(\mathcal{K}')$.

Note that the converse implication of the above theorem is trivial, as the Lebesgue measure on S^{d-1} is invariant under the action of $O(d)$. Note that the function mapping ECC to SECT is injective. Therefore, Theorem 2.10 generalizes to the SECT.

For any $x \in [-a, a]$, define $\delta_x : L^2([-a, a]) \rightarrow \mathbb{R}$ by $\delta_x(f) = f(x)$ for all $f \in L^2([-a, a])$. These functionals, called evaluation functionals, are continuous and, hence, measurable. Then if two embedded simplicial complexes, \mathcal{K} and \mathcal{K}' say, satisfy $\mathcal{K} = \phi(\mathcal{K}')$ for some $\phi \in O(d)$, we get

$$\begin{aligned} \int_{S^{d-1}} \delta_x \circ \text{SECT}(\mathcal{K}) \, d\mu &= \int_{L^2([-a, a])} \delta_x \, d(\text{SECT}(\mathcal{K})_*(\mu)) \\ &= \int_{L^2([-a, a])} \delta_x \, d(\text{SECT}(\mathcal{K}')_*(\mu)) = \int_{S^{d-1}} \delta_x \circ \text{SECT}(\mathcal{K}') \, d\mu \end{aligned} \quad (2.6)$$

for all $x \in [-a, a]$ by the change of variable formula for integrals with measure pushforwards and Theorem 2.10. Hence, the mean of such SECT evaluations, or collections thereof, form a statistic that can be used for distinguishing shapes modulo $O(d)$ actions.

2.2 Temporal shape detection

The SECT transforms a fixed, static shape. We now extend the definition of the SECT to get a rotationally invariant temporal signature of a sequence of shapes.

DEFINITION 2.11 Let $T = [0, c]$ or $T = \{0, 1, \dots, n-1\}$, where $c \in \mathbb{R}$ and $n \in \mathbb{N}$, be a set of time points. Let $\{\mathcal{K}(t)\}_{t \in T}$ be a sequence of geometric simplicial complexes in \mathbb{R}^d and a be such that $\mathcal{K}(t)$ is contained in the ball of radius a around the origin for all $t \in T$. Then the *DETECT* (DEtecting Temporal shape changes with the Euler Characteristic Transform) of this sequence is the transform

$$\text{DETECT}(\{\mathcal{K}(t)\}) : T \rightarrow L^2([-a, a]), \quad t \mapsto \left(x \mapsto \int_{S^{d-1}} \delta_x \circ \text{SEC}_{\mathcal{K}(t)}^v \, dv \right). \quad (2.7)$$

Here, we will use DETECT with $T = \{0, \dots, n-1\}$. In practice, after integrating out the dependence on the direction, DETECT is a continuous function from T to $L^2([-a, a])$. Such functions form an infinite dimensional vector space. Thus, a finite presentation is not generally possible. We, therefore, evaluate DETECT at any fixed $t \in T$ on a finite number of evenly spaced points P in $[-a, a]$. DETECT is then represented approximately by a $|T| \times |P|$ -matrix. In this paper, we apply DETECT to two time-course data sets of organoid boundaries. We will consider the space of such matrices to be endowed with the $\|\cdot\|_2$ -norm.

3. Data sets

We analyse experimental 2D video data as well as synthetic 3D spatio-temporal data.

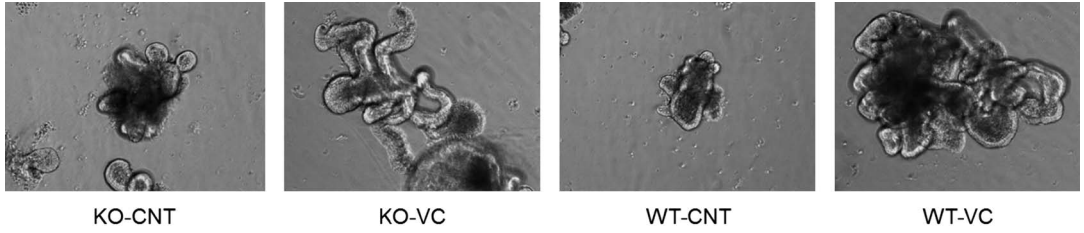


FIG. 3. Phenotype effect of VC treatment to intestinal organoids. Static video snapshots of the final frame. One example is shown for each condition.

3.1 Intestinal organoids from 2D video experiments

We first acquired a set of imaging data derived from time-lapse imaging of mouse small intestine organoids. In total, we have 176 organoids and 320 video frames for each organoid. The data set comprises of 74 wild-type (WT) and 102 p53 knock-out (KO/mutant) genetics organoids. Both groups of organoids further split into untreated (control, abbreviated as CNT) and organoids treated with valproic acid and GSK3 inhibitor CHIR99021 (VC). These organoids have been filmed throughout their growth and the resulting videos have been segmented. After segmentation, we have 100 points summarizing the boundary of each organoid at each video frame (frames starting at beginning of the experiment and are taken every 15 min. henceforth). We discard some videos that for technical reasons have fewer than 320 video frames (as most videos below this threshold still seem to change their morphology at the end of the video). A total of 22 organoids were excluded from the data set in this way to yield 176 organoids.

The organoid boundaries in this data set are, in the 2D video view, close to being perfectly circular in the early stages of the videos across all experimental conditions. This simple geometry is a result of the cellular homogeneity in the early phases. As time progresses, cells proliferate and differentiate. Through proliferation, organoids grow in size and through stem-cell differentiation, the cellular composition of organoids changes. As different cell types have different mechanistic properties and differentiation is not spatially uniform, organoids cease to be spherical. Most notably, they elongate and their boundary buckles, possibly leading to the growth of finger-like protrusions. Such growth behaviour is illustrated by the examples of final video frames presented in Fig. 3.

Each collection of boundary points is transformed into a geometric simplicial complex representing the organoid boundaries, yielding a sequence of geometric simplicial complexes indexed by $t = 0, \dots, 319$ for each organoid. The segmentation algorithm returns 100 points in clockwise ordering for each organoid for each video frame. We connect consecutive points (0-simplices), as well as the first point and the hundredth point, by a line (1-simplex) to obtain a simplicial complex representation. We re-centre each geometric simplicial complex such that the mean of all vertices is the origin. We then compute the radius of the simplicial complex at $t = 0$ (i.e., the largest norm of all vertices after re-centring) and divide all vertices in the sequence of simplicial complexes by that value. We translate and scale the data in this way to simplify it, given its limited size. As a result, the initial size of organoids or any movement throughout time is not considered by any downstream analysis, including DETECT.

3.2 Spatio-temporal synthetic data

We next apply DETECT to a synthetic data set to verify our findings on the experimental data. We use data generated by a model first presented in Yan *et al.* (42). Yan *et al.* construct a mechanistic

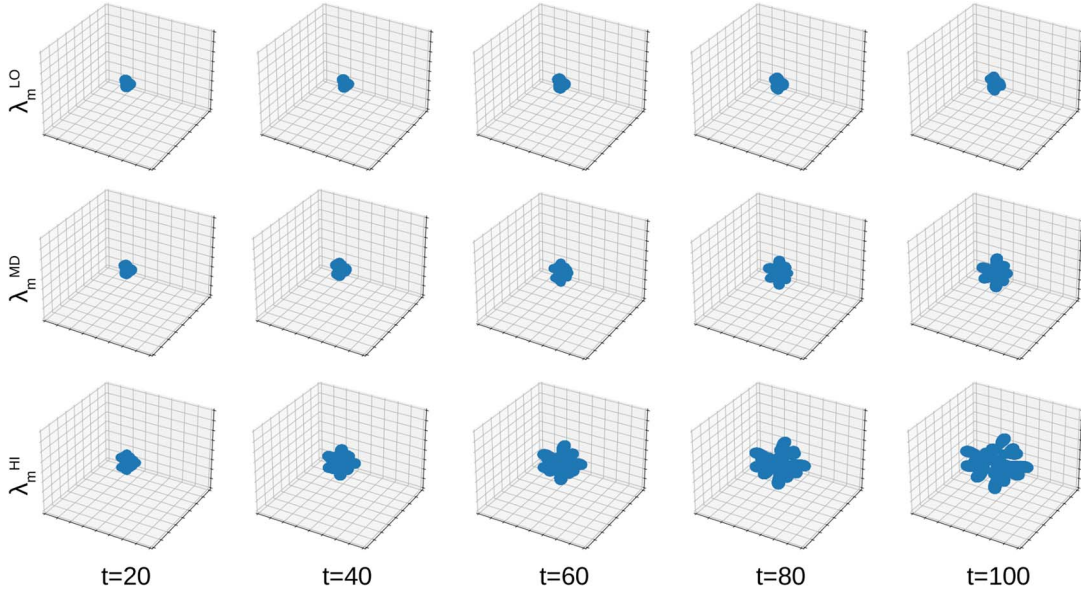


FIG. 4. Spatial shape changes in 3D, synthetic organoids over time. Top row: low (LO) mitosis rate ($\lambda_m^{LO} = 0.35$). Middle row: medium (MD) mitosis rate ($\lambda_m^{MD} = 0.71$). Bottom row: high (HI) mitosis rate ($\lambda_m^{HI} = 1.42$). For all organoids visualized, the rate of dead cell degradation is fixed at the lowest value given in the data set ($\lambda_L = \lambda_L^{LO} = 0.5$).

model describing the 3D growth of colon cancer organoids that contain stem cells (SC), progenitor cells (CP) and terminally differentiated cells (TD). The three cell populations differ in the rates at which they proliferate, differentiate and die. For simplicity, we suppose that the proliferation rates of the stem and progenitor cells are identical and that the differentiated cells do not proliferate, so that $\lambda_m^{SC} = \lambda_m^{CP}$ and $\lambda_m^{TD} = 0$.

We generate a synthetic dataset in which all but two model parameters are fixed at the default, dimensionless values used in (42). We vary two key model parameters: λ_L , the rate at which dead cells are degraded ($\lambda_L = 0.5, 1, 2$), and $\lambda_m^{CS} = \lambda_m^{CT} \equiv \lambda_m$, the rate at which stem and progenitor cells proliferate ($\lambda_m^{SC} = \lambda_m^{CP} = 0.35, 0.71, 1.42$). We perform numerical simulations for all combinations of λ_L and λ_m , to time-course data for nine different synthetic organoids. Illustrative examples of the simulated organoids from this dataset are presented in Fig. 4.

In contrast to the experimental data, the number of (3D) boundary points varies in proportion with the size of the simulated organoids. To ensure computational tractability, we sample 300 boundary points uniformly at random at each time point. Different samples of these 300 points do not lead to any notable perturbations in the downstream analyses (results not presented). To triangulate the boundary surface, we first re-centre the organoid so that the mean of all boundary points is the origin. We then perform a stereographic projection into the xy-plane and perform a Delaunay triangulation and identify those points bordering an infinite area 2-cell. After projecting the finite components of the triangulations back onto the sphere, we add a further point, which is the mean of all points bordering an infinite area cell in the previous step, and insert 1 and 2-cells to fill the north-pole area of our organoid. We credit (1) with this pre-processing procedure. In contrast to the experimental data, re-scaling is not needed as all simulated organoids are identical at $t = 1$.

We perform this pre-processing step to ensure that the resulting geometric simplicial complex is homeomorphic to a sphere (or a union thereof, if an organoid disconnects). This pre-processing is a necessary step to ensure that faithful topological features (e.g., the correct number of components or holes) are present before computing Euler Characteristics. Random rotations of the data ahead of pre-processing lead to negligible differences in ECT and thus suggest this pre-processing does not introduce artificial geometric features.

4. Results

We analyse the shape of organoid boundaries of experimental (2D) and synthetic (3D) data by first building a geometric simplicial complex representation. We then compute the SECT of each organoid at each time point to obtain DETECT. When computing DETECT, we use $a = 6$ for the 2D organoids, $a = 15$ for the 3D organoids (these values are chosen such that all organoids are fully contained within the ball of radius a around the origin at each time). We choose P to be 100 evenly spaced points in $[-a, a]$. We then compute an m -dimensional Nystroem feature embedding ($m = 100$ for the experimental data and $m = 30$ for the smaller synthetic data set; both data sets use a Gaussian kernel with $\lambda = |T| \times |P|$).

4.1 Regressing SECT to classical shape statistics

We first compute the SECT of static images of experimental organoids and demonstrate that the SECT includes information conveyed by classical shape statistics. The classic shape statistics, diameter, the mean and max centroid distances, the equivalent diameter, the major and minor axis lengths and the area of the convex hull quantify geometric properties of a 2D shape. These statistics are widely used and invariant under translation and $O(2)$ actions. As each of these statistics is calculated for a static shape, i.e., for an organoid boundary at a fixed time frame, we compare these statistics to the SECT at fixed time frames.

To compare the aforementioned shape statistics with the SECT, we apply a standard linear regression model. The SECT of each organoid at each time is represented by a Nystroem feature embedding. We project the feature embedding to 50 dimensions using PCA (33). The PCA vectors give the independent variables, while the classic shape statistics listed above are viewed as the dependent variables. We pass the square-root values of the convex hull area to the regression model, as it has a squared relationship with the remaining metrics in the (default) symmetric cases (i.e., scaling a geometric simplicial complex in \mathbb{R}^2 by a scalar x will yield an x^2 -fold change in the convex area but an x -fold change in the six other metrics). An illustration of the main notions is given in Fig. 5. We perform a 50-fold cross-validation for each metric and report mean coefficients of determination and standard deviations of the coefficient of determination in Table 1.

We find that the SECT regresses multiple classical shape statistics with high accuracy. We note that the diameter of an organoid should be proportional to both the major axis length and the maximum centroid distance.

The SECT has high predictive accuracy of equivalent diameters and perimeter and, as a result, can also detect symmetry breaking. The lower accuracy of the minor axis length and convex area suggests that the SECT is more limited in its ability to capture the (mean) size of indentations, compared to detecting size and elongation. We remark that this limitation may be related to segmentation accuracy, and therefore, we consider organoid shapes with known segmentation (i.e., synthetic data) in Section 4.3.

In addition to the above regression analysis, we can decompose the covariance matrix of the aforementioned classical shape statistics by its singular values. We remark that we standardize the data in

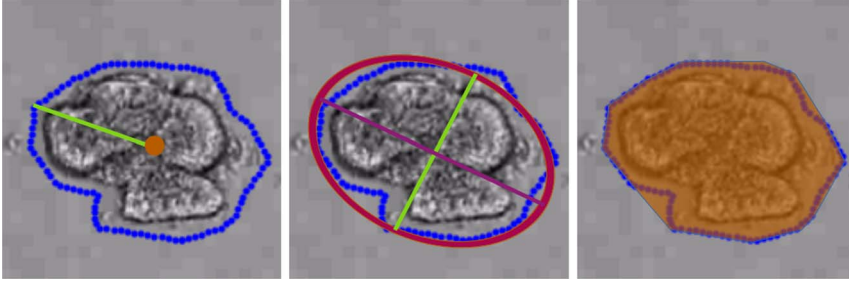


FIG. 5. Left: The red point gives the centroid (point that minimizes mean squared distance to boundary) of the organoid and the green line gives the distance to a boundary point. The maximal and mean lengths of such green lines give the max. and mean centroid distance. Centre: The red ellipse gives the ellipse with the best fit to the boundary. The purple line gives the major axis and the green line the minor axis. Right: the area of the convex hull (convex area) is visualized in opaque red.

TABLE 1 *Mean coefficients of determinations (R^2 -scores) and their standard deviations (std) in a 50-fold cross-validation of a Linear regression in which the Nystroem features of the SECT give the independent variables and the 7 variables above give the dependent variables*

	R^2 -scores	std
equivalent diameter	0.880	0.073
max. centroid distance	0.916	0.041
mean centroid distance	0.894	0.067
major axis length	0.880	0.052
minor axis length	0.630	0.246
perimeter	0.972	0.024
$\sqrt{\text{convex area}}$	0.868	0.105

each feature before computing the covariances. We observe that the first four principal components of the classical shape statistics explain over 90% of the variance in this data set (see Fig. 6). We then perform a canonical correlation analysis (CCA) (40) between these four principal components and the PCA-transformed SECT data. We find that the first four pairs of canonical variables have a perfect correlation score of 1.0 and conclude that the SECT can explain over 90% of the variance in the classical shape statistics on the given data set.

4.2 Classification of organoids

We focus on classifying p53-knock-out organoids into treated and untreated groups. Before training a classifier, we seek to exclude some organoids where the segmentation does not accurately trace the organoid boundary in the video. In these cases, the segmented boundary is significantly larger than the true organoid boundary. Therefore, we exclude organoids whose radius grows by a factor of more than two. For p53-knock-out organoids, this procedure excludes two out of 98 organoids. Of the remaining 96 organoids, 55 are untreated.

We use random forest classification (5) to classify p53-knock-out organoids into untreated and VC-treated experimental groups. Random forest classification then trains an ensemble of decision trees

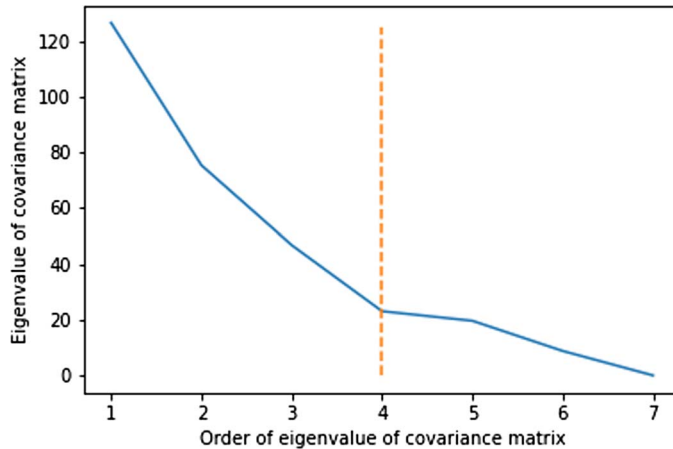


FIG. 6. The eigenvalues (y-axis) of the covariance matrix of the classical shape descriptors in decreasing order (indices of eigenvalues on x-axis). The data was standardized in each component before the covariance matrix was computed. The first four eigenvalues (to the left of the dashed line) account for 90.5% of the variance in the data.

trained on random subsets of the Nystroem-transformed DETECT data. The trees classify data points by majority vote. We use the `scikit-learn` (34) implementation of random forest classification and optimize the hyper-parameters of the maximum tree depth, the minimum number of samples allowed to define a split and the minimum number of samples per tree leaf by cross-validation grid search (`GridSearchCV` in `scikit-learn`). Based on this optimization, maximum tree depth is five, minimum number of samples to define a split is five and minimum number of samples per tree leaf is three. The 5-fold cross-validation for these parameters gives a mean classification accuracy of 68.8% with a standard deviation of 2.7%. A confusion matrix is given in Fig. 7. We remark that increasing the number of Nystroem features from $m = 100$ to $m = 500$ increases the mean accuracy further to 70.0% but also increases the standard deviation to 7.5%. The higher standard deviation suggests that setting $m = 500$ could result in overfitting.

The accuracy of classification results based on DETECT exceeds those based on all classical statistics (e.g., area and perimeter) that give a mean classification accuracy of 60.5% and a standard deviation of 4.8% when we use the pipeline and cross-validation method described above. Classification based on DETECT also exceeds the baseline accuracy associated with guessing, which is 57.4% as there are slightly more untreated than treated organoids in our data set. Further, we have shown that combining DETECT with machine learning can distinguish organoids treated with valproic acid and GSK3 inhibitor based on quantification of their shape dynamics as well as regress out classical shape statistics.

Furthermore, we can compute the feature importances (Gini importances, see e.g., (32)) of the input features. We report the timestamps corresponding to the ten most important features used in our classification in Table 2. We see that the RFC uses features from a variety of different times, illustrating the importance of using time-course data when classifying organoids.

4.3 Distinguishing shape evolution 3D

Finally, we apply our methodology to the synthetic data generated by the model of Yan, Kostorum and Lowengrub (42). As described in Section 3, this data set contains time-courses for 9 organoids. As such,

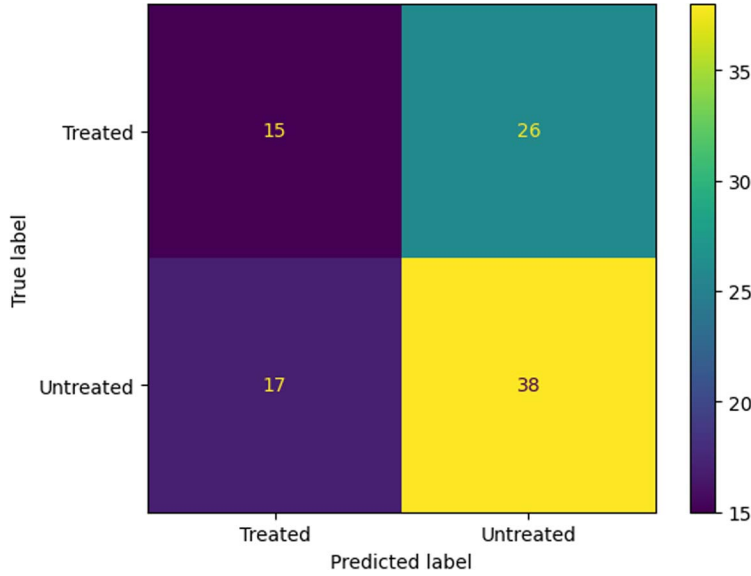


FIG. 7. The confusion matrix of the organoid classification via DETECT and RFC. Based on a 5-fold split of the data set.

TABLE 2 *Feature importances of the Nystroem features passed to an RFC classifier (trained on the whole data set)*

Timestamp of feature	161	278	292	186	237	189	260	261	188	162
Importance (in %)	6.10	6.45	5.44	5.97	7.20	5.05	5.00	4.90	4.71	4.67

it is too small to apply linear regression or random forest classification. Therefore we only report the first two principal components of the Nystroem-transformed DETECT signatures. These outputs demonstrate that our methods generalize well to 3D shapes and that they can identify important structure in the synthetic data, for which we know the ground truth.

This analysis is visualized in Fig. 8 and shows that organoids cluster together by the value of λ_m , the common rate at which the stem and progenitor cells proliferate. This behaviour is consistent with watching the videos for these 9 organoids. We see that organoids with proliferation rates exhibit little growth and virtually no symmetry breaking (see Fig. 4). Organoids with intermediate rates of stem and progenitor cell proliferation ($\lambda_m = \lambda_m^{MD}$) show more growth than organoids with lower proliferation rates ($\lambda_m = \lambda_m^{LO}$) and notable buckling of their boundary (see Fig. 4). Finally, organoids with large rates of stem and progenitor cell proliferation ($\lambda_m = \lambda_m^{HI}$) exhibit the most rapid rate of growth and marked development of protrusions. Indeed, the development of protrusions is so pronounced that several protrusions disconnect from the main organoid at later time points (see Fig. 4).

The principal components in Fig. 8 appear to identify the dominant signals in the synthetic 3D data set. We hypothesize that the first principal component is proportional to the size of the organoids. Similarly, we conjecture that the second principal component corresponds to the geometric complexity of the organoids. In particular, organoids with low rates of stem and progenitor cell proliferation have the lowest geometric complexity while organoids with intermediate rates exhibit significant symmetry

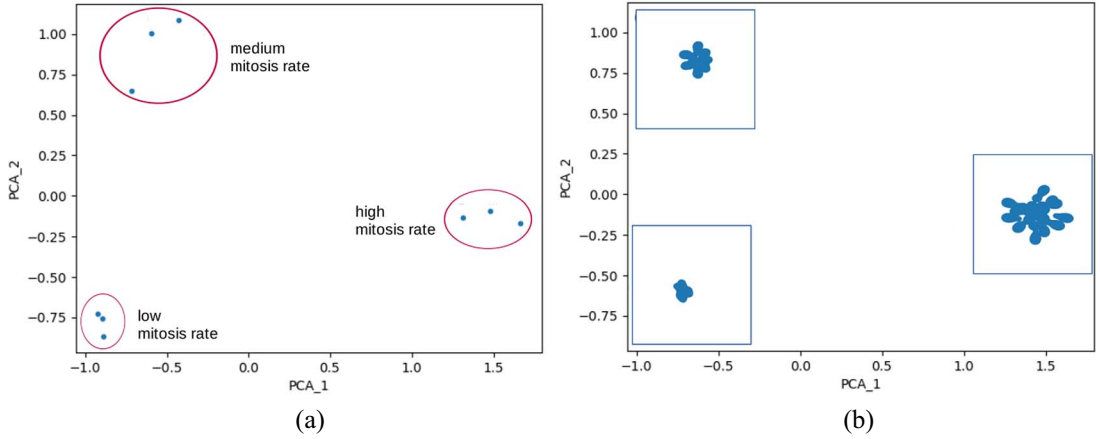


FIG. 8. (A) The first two principal components of the Nystroem-transformed DETECT for the 3D analysis. We see that these Nystroem features of the DETECT signatures cluster by mitosis rate, which is the dominant signal in the data. (B) Example of 2D projections of three organoids with given mitosis rate at the final time-point.

breaking. Organoids with the largest rates of stem and progenitor cell proliferation lie between the two former groups of organoids in terms of their geometric complexity, as their protrusion development is so pronounced that protrusions disconnect. The resulting connected tissues are relatively spherical.

5. Discussion

In this work, we have introduced a new technique from the field of topological data analysis, for detecting temporal shape changes with the Euler Characteristic Transform (DETECT). We have highlighted its utility by studying organoid morphology. We first showed that several classical shape descriptors, including the diameter, the mean and maximum centroid distances, the equivalent diameter, the major and minor axis lengths and the area of the convex hull, can be regressed from the Smooth Euler Characteristic Transform (SECT) with high accuracy. We applied DETECT, with kernel approximation methods and random forests, to a data set of experimental p53 knock-out mouse small intestine organoids, and showed that our approach can distinguish VC-treated organoids from untreated organoids. We remark that this integration with kernel approximations enables larger data sets to be analysed because when kernel methods are used the runtime complexity of ECT can be reduced from being cubic to approximately linear in the number of data points. We highlighted this methodology generalizes to 3D by applying DETECT to an *in silico* data set derived from a continuum model of organoid growth (42). DETECT enables the synthetically generated organoids to be clustered according to mitosis rate (one of the parameters that were varied in the synthetic data).

In the future, we aim to extend our findings—both in 2D and 3D—to data sets of different types of organoids (derived from different organs, with different genetic backgrounds and/or cultured under different conditions). We aim also to study information loss between 3D data and their 2D projections. We will accomplish this by first considering synthetic data generated from 3D mechanistic models (42), which neglect certain biophysical processes (e.g., the effects of gravity, the production of extracellular matrix, mechanical stress) (42). This will enable us to focus on establishing relationships between the DETECT signatures and the values of key model parameters (e.g., cell proliferation and differentiation rates).

We plan to extend this analysis to other types of morphological data that do not have regularized and smooth boundaries. For example, there are no random perturbations in the synthetic data studied here. In practice, data sets analysed by the ECT and its extensions may be noisier than the data sets analysed in this paper. The ECT (and, by extension, its variants, including DETECT) are not stable with respect to small perturbations. Theoretical work studying the stability properties of the ECT and extensions would provide a significant step forward for their application to noisy morphological data sets.

Acknowledgements

We thank Huaming Yan, Anna Konstorum and John Lowengrub for helpful discussions and for providing us with data generated by their model of organoid growth.

Funding

H.A.H. gratefully acknowledges funding from a Royal Society University Research Fellowship. L.M., H.M.B. and H.A.H. are members of the Centre for Topological Data Analysis, which is funded by the EPSRC grant ‘New Approaches to Data Science: Application Driven Topological Data Analysis’ [EP/R018472/1](#). F.Y.Z. was funded by the Ludwig Institute for Cancer Research and an EPSRC Life Sciences Interface Doctoral Training Centre Fellowship, EP/F500394/1. X.Q. was funded by Ludwig Cancer Institute, China Scholarship Council and Oxford University.

REFERENCES

1. AMEZQUITA, E., J., QUIGLEY, M. Y., OPHELDERS, T., LANDIS, J. B., KOENIG, D., MUNCH, E. & CHITWOOD, D. H. (2021) Measuring hidden phenotype: quantifying the shape of barley seeds using the euler characteristic transform. *In Silico Plants*.
2. BEERS, D., HARRINGTON, H. A. & GORIELY, A. (2022) Stability of topological descriptors for neuronal morphology. arXiv preprint arXiv:2211.09058.
3. BOCHE, H., GUILLEMARD, M., KUTYNIOK, G. & PHILIPP, F. (2013) Signal analysis with frame theory and persistent homology. *The Conference of Sampling Theory and Applications, SampTA '13*, vol. 1, p. 1.
4. BREIMAN, L. (2001) Random forests. *Mach. Learn.*, **45**, 5–32.
5. BUSKE, P., PRZYBILLA, J., LOEFFLER, M., SACHS, N., SATO, T., CLEVERS, H. & GALLE, J. (2012) On the biomechanics of stem cell niche formation in the gut—modelling growing organoids. *FEBS J.*, **279**, 3475–3487.
6. BYRNE, H. (2020) Three dimensional biological cultures and organoids. *Interface Focus*, **10**, 20200014.
7. CHUNG, M., K., BUBENIK, P. & KIM, P. T. (2009) Persistence diagrams of cortical surface data. *International Conference on Information Processing in Medical Imaging*. Springer, pp. 386–397.
8. CRAWFORD, L., MONOD, A., CHEN, A. X., MUKHERJEE, S. & RABADAN, R. (2020) Predicting clinical outcomes in glioblastoma: an application of topological and functional data analysis. *J. Am. Stat. Assoc.*, **115**, 1139–1150.
9. CURRY, J., MUKHERJEE, S. & TURNER, K. (2022) How many directions determine a shape and other sufficiency results for two topological transforms. *Trans. Am. Math. Soc. Ser. B*, **9**, 1006–1043.
10. DE SILVA, V. & GHRIST, R. (2007) Coverage in sensor networks via persistent homology. *Algebr. Geom. Topol.*, **7**, 339–358.
11. Delaunay+voronoi on a sphere. <https://www.redblobgames.com/x/1842-delaunay-voronoi-sphere/>. Accessed 02 August 2022. 2024
12. DELGADO-FRIEDRICHS, O., ROBINS, V. & SHEPPARD, A. (2014) Skeletonization and partitioning of digital images using discrete morse theory. *IEEE Trans. Pattern Anal. Mach. Intell.*, **37**, 654–666, 12.

13. DEY, T., K. & MANDAL, S. (2018) Protein classification with improved topological data analysis. *18th International Workshop on Algorithms in Bioinformatics (WABI 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
14. FERAGEN, A., LAUZE, F. & HAUBERG, S. (2015) Geodesic exponential kernels: when curvature and linearity conflict. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 3032–3042.
15. GAMEIRO, M., HIRAOKA, Y., IZUMI, S., KRAMAR, M., MISCHAIKOW, K. & NANDA, V. (2015) A topological measurement of protein compressibility. *Jpn. J. Ind. Appl. Math.*, **32**, 1–17.
16. GHRIST, R., LEVANGER, R. & MAI, H. (2018) Persistent homology and euler integral transforms. *J. Appl. Comput. Topol.*, **2**, 55–60.
17. GIDEA, M. (2017) Topology data analysis of critical transitions in financial networks. arXiv preprint arXiv:1701.06081.
18. GRITTI, N., LE LIM, J., ANLAŞ, K., PANDYA, M., AALDERINK, G., MARTINEZ-ARA, G. & TRIVEDI, V. (2021) Morgana: accessible quantitative analysis of organoids with machine learning. *Development*, **148**, dev199611.
19. HARTUNG, N., MOLLARD, S., BARBOLOSI, D., BENABDALLAH, A., CHAPUISAT, G., HENRY, G., GIACOMETTI, S., ILIADIS, A., CICCOLINI, J., FAIVRE, C., et al. (2014) Mathematical modeling of tumor growth and metastatic spreading: validation in tumor-bearing mice. *Cancer Res.*, **74**, 6397–6407.
20. ISHIHARA, K., MUKHERJEE, A., GROMBERG, E., BRUGUÉS, J., TANAKA, E. M. & JÜLICHER, F. (2022) Topological morphogenesis of neuroepithelial organoids. *Nat. Phys.*, 1–7.
21. JIANG, Q., KURTEK, S. & NEEDHAM, T. (2020) The weighted euler curve transform for shape and image analysis. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 844–845.
22. KANARI, L., DŁOTKO, P., SCOLAMIERO, M., LEVI, R., SHILLCOCK, J., HESS, K. & MARKRAM, H. (2018) A topological representation of branching neuronal morphologies. *Neuroinformatics*, **16**, 3–13.
23. KASSIS, T., HERNANDEZ-GORDILLO, V., LANGER, R. & GRIFFITH, L. G. (2019) Orgaquant: human intestinal organoid localization and quantification using deep convolutional neural networks. *Sci. Rep.*, **9**, 1–7.
24. KIM, S., CHOUNG, S., SUN, R. X., UNG, N., HASHEMI, N., FONG, E. J., LAU, R., SPILLER, E., GASHO, J., FOO, J., et al. (2020) Comparison of cell and organoid-level analysis of patient-derived 3d organoids to evaluate tumor cell growth dynamics and drug response. *SLAS DISCOVERY Adv. Sci. Drug Discov.*, **25**, 744–754.
25. KRAMÁR, M., GOULLET, A., KONDIC, L. & MISCHAIKOW, K. (2013) Persistence of force networks in compressed granular media. *Phys. Rev. E*, **87**, 042207.
26. KRAMÁR, M., GOULLET, A., KONDIC, L. & MISCHAIKOW, K. (2014) Quantifying force networks in particulate systems. *Phys. D: Nonlinear Phenom.*, **283**, 37–55.
27. KRETZSCHMAR, K. & CLEVERS, H. (2016) Organoids: modeling development and the stem cell niche in a dish. *Dev. Cell*, **38**, 590–600.
28. KUMAR, S., MOHRI, M. & TALWALKAR, A. (2012) Sampling methods for the nystrom method. *J. Mach. Learn. Res.*, **13**, 981–1006.
29. LEIBON, G., PAULS, S., ROCKMORE, D. & SAVELL, R. (2008) Topological structures in the equities market network. *Proc. Natl. Acad. Sci.*, **105**, 20589–20594.
30. MARSH, L., DUFRESNE, E., BYRNE, H. M. & HARRINGTON, H. A. (2022) Algebra, geometry and topology of erk kinetics. *Bull. Math. Biol.*, **84**, 1–50.
31. MARTINO, A., RIZZI, A. & MASCIOLI, F. M. F. (2018) Supervised approaches for protein function prediction by topological data analysis. *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–8.
32. NEMBRINI, S., KÖNIG, I. R. & WRIGHT, M. N. (2018) The revival of the gini importance? *Bioinformatics*, **34**, 3711–3718.
33. PEARSON, K. (1901) Liii. on lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dubl. Philos. Mag. J. Sci.*, **2**, 559–572.
34. PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M. & DUCHESNAY, E. (2011) Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

35. PEREA, J. A., DECKARD, A., HAASE, S. B. & HARER, J. (2015) Sw1pers: sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data. *BMC Bioinform.*, **16**, 1–12.
36. TANG, W. S., DA SILVA, G. M., KIRVESLAHTI, H., SKEENS, E., FENG, B., SUDIJONO, T., YANG, K. K., MUKHERJEE, S., RUBENSTEIN, B. & CRAWFORD, L. (2022) A topological data analytic approach for discovering biophysical signatures in protein dynamics. *PLoS Comput. Biol.*, **18**, e1010045.
37. THALHEIM, T., QUAAS, M., HERBERG, M., BRAUMANN, U.-D., KERNER, C., LOEFFLER, M., AUST, G. & GALLE, J. (2018) Linking stem cell function and growth pattern of intestinal organoids. *Dev. Biol.*, **433**, 254–261.
38. TURNER, K., MUKHERJEE, S., BOYER, D. M., BOYER, D. M., MUKHERJEE, S. & CRAWFORD, L. (2014) Persistent homology transform for modeling shapes and surfaces. *Inform. Inf. J. IMA*, **3**, 310–344.
39. WANG, B., SUDIJONO, T., KIRVESLAHTI, H. & GAO, T. (2021) A statistical pipeline for identifying physical features that differentiate classes of 3d shapes. *Ann. Appl. Stat.*, **15**, 638–661.
40. WEGELIN, J., A. (2000) A survey of partial least squares (pls) methods with emphasis on the two-block case.
41. WILLIAMS, C. & SEEGER, M. (2000) Using the nyström method to speed up kernel machines. *Adv. Neural Inf. Proc. Syst.*, **13**.
42. YAN, H., KONSTORUM, A. & LOWENGRUB, J. S. (2018) Three-dimensional spatiotemporal modeling of colon cancer organoids reveals that multimodal control of stem cell self-renewal is a critical determinant of size and shape in early stages of tumor growth. *Bull. Math. Biol.*, **80**, 1404–1433.