Classification of hepatic lesions using the matching metric

Aaron Adcock a,⇑, Daniel Rubin b, Gunnar Carlsson c

a Dept. of Electrical Engineering, 350 Serra Mall, Stanford, CA 94305, United States
b Dept. of Radiology, Richard M. Lucas Center, P285, Stanford, CA 94305, United States
c Dept. of Mathematics, Bldg 380, Stanford, CA 94305, United States

A R T I C L E   I N F O

Article history:
Received 1 October 2012
Accepted 17 October 2013

Keywords:
Medical image processing
Image classification
Persistent homology
Computational topology

A B S T R A C T

In this paper we present a methodology of classifying hepatic (liver) lesions using multidimensional persistent homology, the matching metric (also called the bottleneck distance), and a support vector machine. We present our classification results on a dataset of 132 lesions that have been outlined and annotated by radiologists. We find that topological features are useful in the classification of hepatic lesions. We also find that two-dimensional persistent homology outperforms one-dimensional persistent homology in this application.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Medical imaging technology allows doctors access to portions of the human body which are visually inaccessible to the human eye. Often inspecting these medical images is a labor intensive process performed by diagnostic radiologists. The accuracy of the radiologist is obtained through training and experience [14] but even with extensive training and experience there are variations in interpretations and accuracy among radiologists [4,16]. Despite an increasing emphasis on evidence-based medicine and improved imaging techniques, quantitative ‘gold-standards’ and clear guidelines for a radiologist’s role in quantitative measurements remain elusive [13]. Image processing provides a way of both automating portions of the examination as well as providing standard tools for radiologists to use when reading an image. The qualitative nature of many radiological observations suggests that topological features may be useful in the classification and interpretation of medical images.

In this paper, we explore automatic classification methods of computed tomography (CT) scans of hepatic (liver) lesions. We have a dataset of CT scans of 132 hepatic lesions along with an outline, diagnosis and semantic descriptors of the lesion provided by a radiologist. There are nine lesion types represented in the data, with the vast majority of the lesions (90 lesions) evenly split between cysts and metastases, followed by hemangiomas (18 lesions), hepatocellular carcinomas (HCC, 11 lesions), focal nodules (5 lesions), abscesses (3 lesions), neuroendocrine neoplasms (NeN, 3 lesions), a single laceration and a single fat deposit (see Figs. 1 and 2).

It has been demonstrated that semantic features are useful for classification in hepatic lesions [14]. This indicates that visually identifiable structures exist within the lesions, but it has been difficult finding quantitative methods of defining these structures. For example, consider the six images in Figs. 3 and 4. The first three show the abscesses contained in our dataset. The second three show hemangiomas (deformations of blood vessels). The abscesses present what is called ‘cluster of grapes’ morphology. But the arrangements of this structure are very different in each lesion. Similarly, the hemangiomas show the characteristic large dark central region with dense white regions on the outer edge of the lesion. Yet, the hemangiomas lack a rotational orientation, different numbers of the two region types exist and the formations vary in size and shape. The qualitative nature of these observations has made it difficult to find quantitative measures of the structures.

1.1. Prior work

As mentioned above, semantic features have been one successful method for classifying the liver lesions [14]. This has led to preliminary investigations into using quantitative features to predict semantic features, which can be used to classify the lesions [12]. Additionally, computational features have shown some success in directly classifying liver lesions [12,14,18]. Most of these studies use a large number of various types of features (intensity histograms, wavelets, boundary features, etc.) to classify the lesions. Shape descriptors for liver lesions have also been investigated and found to work well in retrieving similar lesions [17].
Persistent homology [5,19] is an approach to extending the notion of shape to point clouds or finite metric spaces, which has been developed over the last 10–15 years. It has two directions of application, one which gives understanding of the overall organization of data sets (see e.g. [1,6,15]), and a second which applies to data sets consisting of data points which themselves have complex structure, such as databases of images or of chemical structures. This paper approaches radiological images from the second point of view. Persistent homology depends on a family of simplicial complexes parametrized by a real variable. In many applications, including the ones in the first direction, this parameter is simply a scale variable, which measures distances between points. In applications in the second direction, however, one uses families depending on other parameters, and often uses multiple persistence invariants in conjunction with each other to obtain useful information (see [8,10,11]). In particular, the authors of [11] use size functions to measure symmetry in classifying potential melanocytic lesions. In this paper we will find an appropriate set of parameters and functions of those parameters for classification problems arising in the radiology of liver lesions.

1.2. Our work

In this paper we present a framework for computing persistent homology on images. This framework is flexible and allows the user to tailor the filtrations to the application. We demonstrate this by applying this framework to classifying our set of hepatic lesions and using the bottleneck distance to compare lesions. We find that our method gets comparable results to existing methods and our results demonstrate the possibilities of tailoring multidimensional

Fig. 1. Abdominal CT scan.

Fig. 2. Lesion diagnoses.

Fig. 3. Abscesses.
persistence to a specific application. Additionally, since we use an 'off-the-shelf' implementation of the support vector machine to perform the final classification, our results demonstrate that this framework can be easily integrated with existing classification techniques.

2. Theory

The work presented in this paper is an application of computational topology to medical image processing. As such, the theoretical material will be presented in a compact, informal manner following [8,10]. Those interested in the details behind the theory are encouraged to read the associated references.

2.1. Filtrations and simplicial complexes

Let \( V \) be a set of points. A \( k \)-simplex \( S \subseteq V \) of cardinality \( k+1 \). A geometric realization of a simplex consists of the convex hull of \( k+1 \) affinely independent points in \( \mathbb{R}^d, d \geq k \). A 0-simplex is a point or vertex, a 1-simplex is an edge, a 2-simplex is a triangle, and a 3-simplex is a tetrahedron. Higher dimensional simplices are difficult to visualize and thus less familiar to our experience. A simplicial complex on \( V \) is a set \( K \) of simplices on \( V \) such that the following holds: If the simplex \( \sigma \in K \), and the simplex \( \tau \subset \sigma \) then \( \tau \in K \). A subcomplex of \( K \) is a simplicial complex \( L \subseteq K \). A filtration of a complex \( K \) is a nested sequence of complexes \( 0 = K_{\text{min}} \subseteq K_{\text{min}+1} \subseteq \ldots \subseteq K_{\text{max}} = K \). We call \( K \) a filtered complex. We will often use the term complex to refer to both the complex itself and a geometric realization of the complex.

2.2. Persistent homology

Homology is an algebraic invariant that counts some of the topological invariants of the geometric realization \( X \) of a complex in terms of its Betti numbers, \( \beta_i \). Specifically, \( \beta_0 \) counts the number of connected components of \( X \). \( \beta_1 \) counts the number of tunnels through \( X \), and \( \beta_2 \) counts the number of voids in \( X \) (three-dimensional empty space enclosed by a two-dimensional surface).

Consider the filtered complex, a nested sequence of complexes, mentioned above. We can track the topological changes that occur in this sequence via persistent homology, an algebraic invariant which tracks the birth (appearance) and death (disappearance) of topological attributes (Betti numbers) during the evolution of this sequence. The birth \((a)\) and death \((b)\) of a feature can be represented using the interval \([a,b]\) \( \in \mathbb{R} \). Counting the number of features in existence at a point \( t \) (intervals which contain \( t \)) in the persistent homology of the filtered complex gives the homology of \( K_t \). Generally, the longer a feature's lifetime, the more important the feature is considered to be. Features can have infinite lifetimes if they appear in all complexes after a given point in the sequence.

These concepts can be extended to multidimensional filtrations [7]. For example, in a two-dimensional filtration we have a sequence of simplicial complexes such that \( K_{i_1,i_2} \subseteq K_{i_1,i_2} \) if \( i_1 \leq J_1 \) or \( i_2 \leq J_2 \). The topological features in the two-dimensional case are more complicated than their one-dimensional counterparts but for the purposes of this application, can be thought of as existing across two-dimensional 'sheets' rather than bars. Except where noted, it can be assumed we are referring to one-dimensional filtrations.

2.3. A metric on barcodes

By computing the persistent homology of a filtered complex, we obtain a descriptor of the complex in the form of a finite multi-set of intervals, called a barcode. Thus, the barcode is useful both as a data structure for storing the results of the computation of the persistent homology of a filtered complex and as a visual representation of the persistent homology. A quasi-metric \( D \), which we define as a metric which can take infinite values, can be defined over the collection of all barcodes allowing us to compare complexes using this quasi-metric on the space of barcodes. We follow [10] in defining \( D \).

Let \( B_1 \) and \( B_2 \) be two barcodes or finite multi-sets of intervals. For two intervals, \( I \) and \( J \), we define their dissimilarity \( \delta(I,J) \) to be their symmetric difference: \( \delta(I,J) = \mu(I \cup J - I \cap J) \), where \( \mu \) denotes the one-dimensional measure. Note that \( I \cup J \) can be infinite intervals and consequently \( \delta(I,J) \) can be infinite. A matching \( M \) on \( B_1 \) and \( B_2 \) is a set \( M(B_1,B_2) \subseteq B_1 \times B_2 = \{(I,J)|I \in B_1, J \in B_2\} \), where each interval occurs in at most one pair \((I,J)\). Note that in general, \( M \) will not have matched all intervals from \( B_1 \) and \( B_2 \). Let \( N \) be the set of unmatched intervals. We can now define \( D_M(B_1,B_2) \), or the distance relative to \( M \), as

\[
D_M(B_1,B_2) = \sum_{(I,J) \in M} \delta(I,J) + \sum_{L \in N} \mu(L)
\]

We define the quasi-metric \( D(B_1,B_2) \) as the best possible matching between \( B_1,B_2 \):

\[
D(B_1,B_2) = \min_M D_M(B_1,B_2)
\]

This problem can be recast maximum weight bipartite matching problem and solved using the Hungarian algorithm. See [10] for details. We will abuse nomenclature slightly and refer to \( D \) as the matching metric.

3. Calculations

Each point in our dataset is an image, a two-dimensional collection of pixels (grayscale or intensity values) laid out on a grid with a set of contiguous pixels marked as lesion tissue. To use computational topology to analyze each image, we need a method of forming a filtered complex from an image. We can then use the theory outlined in Section 2 to create a barcode for each image and then use the matching metric to compare various images.
3.1. Forming a simplicial complex from an image

Given a two-dimensional image $I$, we begin with the empty complex $K$. We then assign a vertex to each pixel in $I$ and add each of these vertices (0-simplices) to $K$. We then form 1-simplices from these vertices if the associated pixels are adjacent in $I$ (we treat diagonal pixels as adjacent). We then add 2-simplices on the vertices where 3 pixels are mutually adjacent. This forms a very regular simplicial complex, a mesh, with the only variations between images being the boundary shape of $I$. As we are interested only in the hepatic lesions identified in the image by the radiologist, we define $I$ to be the region of pixels contained within the lesion outline, plus a border of healthy tissue around the edge of the lesion. We set the width of this border to 5 pixels. We keep this border because it is useful to have some healthy tissue included in the filtered complex for comparison with the lesion tissue.

3.2. Filtered filtrations

After constructing the simplicial complex $K$ on the image $I$, we now want to define a filtration on $K$. A natural approach is to assign a value to each vertex. We can represent this as a function $f : V \rightarrow \mathbb{R}$, where $V$ is the vertex set of $K$. Let $f_{\text{min}}$ and $f_{\text{max}}$ denote the minimum and maximum values obtained by $f$ on $V$. We construct $K_f$ by including any simplex $S \in K$ with the property $\forall \nu \in S, f(\nu) \leq i$.

\[ K_i = \{ S \in K : \forall \nu \in S, f(\nu) \leq i \} \quad (1) \]

Intuitively, $f(\nu)$ represents the point at which $\nu$ enters the filtration and $\max_{\nu \in S} f(\nu)$ determines the point at which a simplex $S \in K$ enters the filtration. Let $i_1 < i_2 < \ldots$ be an increasing sequence of positive values. Then we have the filtered complex $K_{f_{\text{min}}} \subseteq K_{f_{\text{min}}+i_1} \subseteq K_{f_{\text{min}}+i_2} \subseteq \ldots \subseteq K_{f_{\text{max}}} = K$. Notice that if we reverse the inequality in Eq. (1) we get an equally valid filtration, $K_{f_{\text{max}}} \subseteq K_{f_{\text{max}}-i_1} \subseteq \ldots \subseteq K_{f_{\text{min}}} = K$. We will refer to these as the increasing and decreasing filtrations.

As each vertex is associated with a pixel in the original image, it is natural to use the pixel intensity (i.e., the grayscale value of each pixel) to assign a value to each vertex. This forms the basis of what we will call the intensity filtration. A toy example of the increasing intensity filtration is shown in Fig. 5a. The colors represent the point in the filtration when the vertices and edges are added (we do not shade triangles for aesthetic reasons).

We define an additional filtration by associating the distance from the lesion border, as given by the radiologist, to each pixel. We call this the border filtration. The increasing border filtration produces an 'annulus' which grows until it fills the lesion. The decreasing filtration produces a misshapen 'disc' which expands from the center of the lesion. While this is clearly not topologically useful for classification, in practice the combination of the border filtration with the intensity filtration gives better classification results than using the intensity filtration alone.

To simplify the computational and theoretical difficulties encountered with two-dimensional filtrations, we use one-dimensional filtration slices to approximate the two-dimensional filtrations. This simple sampling method is effective in our application, but the reader interested in a more rigorous treatment of dimensional reductions of multidimensional persistence is referred to [2,3]. Let $K_{f}^i$ represent the border filtration and $K_{f}^p$ represent the pixel intensity filtration. We divide the range of the border-filtration into 20 equally spaced slices. At each slice $i$, we use the intensity filtration to compute the persistent homology of the subcomplex $K_{f}^i$. This gives rise to the filtered complex $K_{f_{\text{max}}} = K_{f}^0 \cap K_{f}^1 \cap \ldots \subseteq \subseteq K_{f_{\text{min}}} = K_{f}^0 \cap K_{f}^1 \cap \ldots = K_{f}^p$.

We can treat each of these one-dimensional barcodes as a different measurement on the lesion. The options of the increasing or decreasing filtration on the border and intensity functions, as well as the $\beta_0$ and $\beta_1$ barcodes gives eight barcodes at each slice, yielding a total of 160 barcodes computed on each lesion.

To account for differences in the pixel scaling, whether due to image formatting or differing CT scanners, we normalize the pixel range from zero to 1. We stop infinite barcodes at 1.1 so that a differing number of infinite bars does not immediately separate two lesions.

3.3. Feature generation and machine learning

To make use of existing machine learning techniques, it is necessary to provide a vector of measurements for each lesion. Using the barcode distance, we can create a vector of relative measurements by computing the matching distance between the barcodes generated on each lesion. In other words, we use the entire set of 132 images as the comparison set to generate a $132 \times 132$ square matrix $D$ per barcode. Each entry $D_{ij}$ is the distance between the barcode on the lesion $i$ and the corresponding barcode on lesion $j$. Since we have multiple barcodes for each lesion (160 using the sampled two-dimensional filtration), we compute the distance

![Fig. 5. Constructing an increasing 1D-filtration on an image.](image-url)
between all of the corresponding barcodes. This yields many of the 
square distance matrices mentioned above. We choose to sum all 
of these matrices to create a single matrix. We can then use the col-
umns of this matrix as feature vectors. We do this even when 
restricting ourselves to a smaller subset of lesions and simply keep 
a subset of the same columns (still containing 132 entries). This 
allows us to obtain information even from the lesion type sets that 
are too small for classification.

This vector can then be used in traditional machine learning 
algorithms. We choose to use a common implementation of the 
support vector machine (SVM) called LibSVM to test the classifica-
tion accuracy on various subsets of the data with the one-
dimensional and two-dimensional filtrations [9]. We present the 
results using a Gaussian kernel (also known as the radial basis 
function) with the SVM and an explicit definition of this kernel is 
given in the next section.

4. Results

For the purpose of building intuition, we used classical multi-
dimensional scaling (CMDS) on the distance matrix, using the above 
feature vectors as columns, to produce 2D and 3D visualizations of 
the lesions, shown in Fig. 6. Note that the axes simply give the 
coordinates of the embedding given by CMDS and that the vertical 
axis in Fig. 6b is pointed at the reader in Fig. 6a.

We use a SVM and leave one out cross-validation (LOOCV) to 
test the efficacy of our methods. Because the dataset available to 
us is very unbalanced, see Section 1, we present results from four 
different subsets of the data. The first is the full dataset. In the sec-
ond subset (HcHeCM), we use the HCCs, the hemangiomas, the 
cysts, and the metastases. In the third dataset (HeCM) we test on 
the hemangiomas, cysts, and metastases. This data set is most 
comparable to the sets used in previous classification works [14].

![Fig. 6. Visualizations of topological features.](image-url)
In the final dataset (CM) we remove the hemangiomas, leaving only the cysts and metastases (see Table 1).

We used the Gaussian kernel $e^{-|u-v|^2}$ where $u$ and $v$ are two feature vectors, in combination with the SVM. We performed an exponential parameter sweep to find reasonable values for the parameter $\sigma$ and the SVM cost parameter $C$ for each data set. In Table 2 we show the classification rates of each lesion type in the HeCM dataset.

Upon examination of the lesions which were misclassified, we noticed that many of the lesions were significantly larger than the median lesion area (1285.5 pixels) of the dataset. Taking HeCM, we performed the same analysis as above, but removed lesions with various pixel areas. The results are summarized in Table 3.

5. Discussion

These results demonstrate the power of combining topology and geometry via persistence. In this case, it is clear that using a radial geometry (via the border filtration) significantly improves the classifying power of barcodes, especially in the case of hemangiomas, which are characterized by large homogenous regions on the periphery of the lesion. In other words, we were able to develop a 'soft' geometric feature using persistent homology and medical expertise that was useful for classification. The ability to capture qualitative observations makes this technique a powerful alternative to classic learning approaches.

A large portion of the misclassifications are due to the difficulties in normalizing for the number of pixels present in the lesion image. A large lesion, has many more potential topological generators than a smaller lesion. Since the matching metric involves a matching of bars, a larger lesion will be a greater distance from a smaller lesion, even if the tissue structure is the same, because of the unmatched bars. A method of accounting for the differences in lesion size would improve our results considerably. In addition to this, our results are potentially sensitive to the lesions available for comparison. A different comparison set (or even a synthetic comparison set) could improve or degrade the results. Addressing both these issues are potential future directions of research.

This procedure is flexible enough to be used in a variety of contexts. Any method of assigning values to each pixel could be used as a filtration to generate a barcode. Filtering could be done on the image or on the barcodes (for example, removing bars of small length). If a different geometric filter is called for in an application, it can be easily accommodated by our methods; though different techniques are required to move beyond two-dimensional filtrations.

Nevertheless, the current results are comparable to more traditional feature based classification methods [14,18]. Since topological methods are novel for the classification of hepatic lesions, a mixture of classical and topological techniques could prove to be a powerful classifier. As our use of the SVM demonstrates, topological techniques would be fairly easy to integrate with the existing methods mentioned above. As such, we believe that combining topological features with more commonly used approaches could be a fruitful area for future research in image classification.

6. Conclusion

We have implemented and tested a methodology for the classification of hepatic lesions. Using multidimensional persistent homology and a support vector machine, we demonstrated the ability of multidimensional persistence to combine different features of interest for improved results over a one-dimensional filtration. Our computational framework can be used in a variety of image classification problems outside of lesion classification and can be tailored to the specific application by changing the filtrations used on the image. We achieved comparable results to the traditional classification methods of hepatic lesions and, because our methods are topologically based, this makes them good candidates for integration with classical non-topologically based image processing techniques.

References

[1] H. Adams, G. Carlsson, On the non-linear statistics of range image patches, SIAM J. Imag. Sci. vol. 1, no. 2, pp. 110–117.
[2] F. Cagliari, B. Di Fabio, M. Ferri, One-dimensional reduction of multidimensional persistent homology, Proc. Am. Math. Soc. 138 (2010) 3003–3017.
[3] S. Biasotti, A. Cerri, P. Frosini, D. Giorgi, C. Landi, Multidimensional size functions for shape comparison, J. Mathemat. Imag. Vision 32 (2) (2008) 161–179.
[4] S.G. Armato, M.F. McNitt-Gray, A. Reeves, G. McLennan, D.R. Rubin, Automated retrieval of CT images of liver lesions on the basis of image classification, Comput. Med. Imag. and Graph. 31 (2007) 91–103.
[5] G. Carlsson, Topology and data, Bull. Am. Math. Soc. 46 (2009) 255–308.
[6] G. Carlsson, T. Ishkhanov, V. de Silva, A. Zomorodian, On the local behavior of spaces of natural images, Int. J. Comput. Vision 76 (2008) 1–12.
[7] G. Carlsson, A. Zomorodian, Theory of multidimensional persistence, Discr. Comput. Geom. 42 (1) (2009).
[8] G. Carlsson, A. Zomorodian, A. Collins, L. Guibas, Persistence barcodes for shapes, Int. J. Shape Model. (11) (2005) 149–187.
[9] Chih-Chung Chang, Chih-Jen Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (2011) 27:1–27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[10] A. Collins, A. Zomorodian, G. Carlsson, L. Guibas, A barcode shape descriptor for curve point cloud data, Comput. Graph. 28 (2004) 881–894.
[11] Massimo Ferri, Ignazio Stanganelli, Size Functions for the Morphological Analysis of Melanocytic Lesions, Int. J. Biomed. Imag. 2010 (2010), Article ID 621357.
[12] F. Gimenez, Xu Jiajing, Liu Yi, T.T. Liu, C.F. Beaulieu, D.L. Rubin, S. Napel, On the Feasibility of Predicting Radiological Observations from Computational Imaging Features of Liver Lesions in CT Scans, in: 2011 First IEEE International Conference on Healthcare Informatics and Systems Biology, pp. 346–350, 26–29 July, 2011.
[13] T.A. Jaffe, N.W. Wickersham, D.C. Sullivan, Quantitative imaging in oncology patients: Part 1, radiology practice patterns at major U.S. cancer centers, AJR 193 (2010) 101–106.
[14] Sandy A. Napel, Christopher F. Beaulieu, Cesar Rodriguez, Jinyu Cui, Jiajing Xu, Ankur Gupta, Daniel Korenblum, MS, Hayit Greenspan, Yongjun Ma, Daniel L. Rubin, Automated retrieval of CT images of liver lesions on the basis of image classification.

Table 1
SVM Classification accuracies for 1D and 2D filtrations.

| Filtration | Full (%) | HeCM (%) | HeCM (%) | CM (%) |
|-----------|----------|----------|----------|--------|
| 1D (intensity) | 55.30 | 59.66 | 63.89 | 80.00 |
| 2D | 66.67 | 72.27 | 80.56 | 85.56 |

Table 2
HeCM % classification accuracy by lesion type.

| Filtration | % of HeCM | % of Heman | % of Cysts | % of Metas. |
|-----------|-----------|------------|------------|-------------|
| 1D | 62.89 | 27.78 | 77.78 | 64.44 |
| 2D | 80.85 | 72.22 | 88.89 | 75.56 |

Table 3
Classification by lesion size of HeCM.

| Lesion size by area | % Accu. | # of Heman | # of Cysts | # of Metas. |
|---------------------|---------|------------|------------|-------------|
| All | 80.56 | 18 | 45 | 45 |
| <1:000 px | 83.50 | 18 | 42 | 43 |
| <5000 px | 86.96 | 16 | 39 | 37 |
| <2500 px | 86.25 | 14 | 32 | 34 |
| <1250 px | 91.53 | 8 | 28 | 23 |
similarity: method and preliminary results, Radiology vol. 256, no. 1, 2010, pp. 243–252.

[15] J. Perea, G. Carlsson, A Klein-Bottle-based dictionary for texture representation, In Preparation.

[16] P.J. Robinson, Radiology’s Achilles heel: error and variation in the interpretation of the Röntgen image, Br. J. Radiol. 70 (839) (1997) 1085–1098.

[17] J. Xu, J. Faruque, C.F. Beaulieu, D.L. Rubin, S. Napel, A comprehensive descriptor of shape: method and application to content-based retrieval of similar appearing lesions in medical images, J. Digit. Imag. (2011).

[18] C.G. Zhao, H.Y. Cheng, Y.L. Huo, T.G. Zhuang, Liver CT-image retrieval based on Gabor texture, in: Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE, vol. 1, no., pp.1491–1494, 1–5 September, 2004.

[19] A. Zomorodian, G. Carlsson, Computing persistent homology, Discr. Comput. Geomet. 33 (2) (2005) 247–274.

Aaron Adcock is a PhD candidate in the Department of Electrical Engineering at Stanford University and is advised by Gunnar Carlsson and Michael Mahoney. His research interests include applications of computational topology, the analysis of complex networks, and data mining. His current projects include developing new techniques in applying computational topology to medical images, integrating computational topology with existing machine learning techniques, and investigating the structure of complex networks using tree-like structures.

Daniel Rubin is Assistant Professor in the Department of Radiology at Stanford University. Work in the lab lies at the intersection of biomedical informatics and imaging science. His NIH-funded research program focuses on developing informatics methods of knowledge representation, natural language processing, and decision support to improve the quality and consistency of radiology practice. Major projects include (1) developing methods to extract information and meaning from images for data mining, (2) developing statistical natural language processing methods to extract and summarize information in radiology reports and published articles, (3) resources to integrate images with related clinical and molecular data to discover novel image biomarkers of disease, and (4) translating these methods into practice by creating decision support applications that relate radiology findings to diagnoses and that will improve diagnostic accuracy and clinical effectiveness.

Gunnar Carlsson is the Swindells Professor of Mathematics at Stanford University and is a co-founder of Ayasdi Inc. Gunnar has a Ph.D. in Mathematics from Stanford University. In the last ten years, he has been involved in adapting topological techniques to data analysis, under NSF funding and as the lead PI on the DARPA Topological Data Analysis project from 2005 to 2010. He is the lead organizer of the ATMCS conferences, and serves as an editor of several Mathematics journals.