Novel integration of topological data analysis into a multimodal machine learning model to predict follicular carcinoma on ultrasound

Andrew M Thomas, PhD[1], Ann C Lin, MD[3], Grace Deng, PhD[2], Yuchen Xu, PhD[2], Gustavo Fernandez Ranvier, MD, PhD[3], Aida Taye, MD[3], Randall Owen, MD[3], David S Matteson, PhD[2], Denise Lee, MD[3]

[1]Department of Statistics and Actuarial Science, University of Iowa
[2]Department of Statistics and Data Science, Cornell University
[3]Department of Surgery, Icahn School of Medicine at Mount Sinai

This manuscript was a podium presentation at the 2023 American Thyroid Association Meeting in Washington DC, September 27–October 1, 2023.

Keywords: topological data analysis, follicular carcinoma, machine learning, multimodal, thyroid nodule, ultrasound

Corresponding Author:
Denise Lee, MD
Division of Surgical Oncology and Endocrine Surgery
Department of Surgery
Icahn School of Medicine at Mount Sinai
New York, NY 10029
Telephone: 212-241-2891
Fax: 212-241-1572
denise.lee@mountsinai.org

Abstract

Background:
Multiple ultrasound (US) risk stratifications systems (RSS) have been developed to estimate malignancy risk in thyroid nodules and recommend the need for fine needle aspiration biopsy. However, sonographic risk patterns identified in established RSSs may not accurately stratify follicular carcinoma from adenoma, which share many similar US characteristics. Quantitative medical imaging analysis aims to extract high-dimensional textural features from tumor phenotypes imperceptible to the human eye. The purpose of this study is to investigate the performance of a multimodal machine learning model utilizing radiomics and topological data analysis (TDA) to predict malignancy in follicular thyroid neoplasms on ultrasonography.

Methods:
This is a retrospective study of patients who underwent thyroidectomy with pathology confirmed follicular adenoma or carcinoma at a single academic medical center between 2010–2022. The nodule of interest on pre-operative ultrasound was annotated and masked to only include pixels. Images were scaled to maintain aspect ratio and ensure similar image resolution. Features derived from radiomics and TDA were calculated. High-dimensional radiomics and TDA features were projected onto their first two principal components respectively. Logistic regression with L2 penalty was used to predict malignancy. Classifier performance was evaluated using leave-one-out cross-validation and area under the curve (AUC).

Results:
Patients with follicular adenomas (n=7) and follicular carcinomas (n=11) with available imaging were included. 910 radiomics features were extracted for each image. 180 topological features from the height filtration, mean and variance of pixel intensities, the aspect ratio of images and two additional persistence statistics were derived from each image. The models achieved an AUC of 0.68 (radiomics only) and 0.88 (radiomics and TDA). The subsampling bootstrap was used to assess the confidence of these estimates.

Conclusions:
We demonstrate that inclusion of topological features yields strong improvement over radiomics-based features alone in the prediction of follicular carcinoma on ultrasound. Despite low volume data, the TDA features of the height filtration with its connection to the persistent homology transform (PHT), explicitly capture shape information that likely augments performance of the multimodal machine learning model. This approach suggests that a quantitative based US RSS may contribute to the preoperative prediction of follicular carcinoma.

Introduction

Thyroid nodules are common in the general population, and standard evaluation begins with ultrasound imaging. The decision to proceed with fine needle aspiration biopsy (FNAB) is based on nodule size and the presence of suspicious sonographic features.[1] Multiple risk stratification systems (RSS), such as the Thyroid Imaging Reporting and Data System (TI-RADS) and American Thyroid Association (ATA) sonographic pattern guidelines, have been developed to systematically evaluate thyroid ultrasounds and to estimate malignancy risk in thyroid nodules.[2,3] However, there continues to be a lack of standardization in the use of these various RSSs for thyroid nodules, as well as in the clinical interpretation of their assigned values.[4-6] Additionally, distinguishing benign follicular adenomas from follicular carcinoma has posed a particular diagnostic dilemma, as they share many similar sonographic characteristics and may not be accurately stratified on RSSs.[7,8] Since follicular thyroid neoplasms cannot be distinguished on cytology, final diagnosis can only be made after surgical excision with microscopic examination for capsular or vascular invasion.[9,10] An estimated 80-90% of follicular thyroid neoplasms are benign, resulting in a significant number of surgeries performed for benign disease.[11]

In recent years, the paradigm of work up of thyroid nodules has shifted to a more conservative approach to prevent unnecessary biopsies, and ultimately unnecessary surgeries. As such, the implementation of innovative techniques to improve risk stratification of follicular thyroid neoplasms based solely on non-invasive tools, such as ultrasound, are of particular interest. Quantitative medical imaging analysis is the process of extracting large numbers of high-dimensional textural features from tomographic images, which can then be used to inform decision support tools. Radiomics specifically refers to extraction of pixel-level features within a region of interest of the image, converting the image into high-dimensional data.[12] Alternatively, topological data analysis (TDA) specifically utilizes tools from algebraic topology, such as persistent homology and persistence diagrams[13], to examine structures on multiple scales, focusing on the shape of data.[14,15]

While these techniques can be applied to many clinical scenarios, this area of research has had particular focus in oncology, as these processes have potential to identify and refine tumor phenotypes on imaging beyond the scope of human analysis. Previous studies have applied radiomics to thyroid ultrasound, with promising results in predicting malignancy in thyroid nodules in comparison to radiologist review using RSSs.[16,17] And while TDA has been applied to the diagnosis of melanoma, breast, brain, colorectal, and lung cancer, there is a dearth of research in the utilization of TDA in thyroid imaging[14,18].

These two quantitative imaging analysis techniques have potential to enhance the diagnostic ability of ultrasound images in distinguishing follicular thyroid neoplasms by capturing features that are imperceptible to the human eye. The objective of this study is to investigate the performance of a multimodal machine learning (ML) model utilizing both radiomics and TDA to predict malignancy in follicular thyroid neoplasms on ultrasound images.

Methods

Data Collection

This study was approved by the Icahn School of Medicine Institutional Review Board. A cohort of adult patients (N=908) who underwent thyroid surgery at our tertiary referral institution between 2010 and 2022 were identified. Of these, patients with diagnosis of follicular carcinoma or follicular adenoma on final pathology were reviewed (N=82). Exclusion criteria were as follows: history of prior thyroid cancer (N=5), multiple cancer diagnoses in final pathology (N=2), diagnosis of Hurthle cell adenoma or carcinoma (N=7), no pre-operative thyroid ultrasound images available (N=50). A total of 18 patients

remained for inclusion in this study, with N=7 patients in follicular adenoma cohort and N=11 in follicular carcinoma.

All available pre-operative thyroid ultrasound images for each patient were downloaded from the Picture Archiving and Communications System (PACS) for analysis. Each nodule of interest was identified by cross-referencing the images with cytopathology and final pathology reports. The images including the nodule in view were selected and annotated to indicate region of interest (ROI) by an endocrine surgery fellow (A.C.L). For every patient, a representative image was chosen, and downscaled to the minimum image resolution so that each image would have roughly the same number of pixels.

Topological Data Analysis

Persistent homology[13] is a mathematical tool that can be used to detect quantitative shape information such as connected components (0-dimensional), holes/loops (1-dimensional), voids (2-dimensional), etc.—at different scales. Two-dimensional images only contain 0- and 1-dimensional topological features. To calculate the 0- or 1-dimensional persistent homology for a greyscale image X, we first require a *filtration function* I, on the image X. We then examine the shape of the binary images I_t, which consist of all pixels p in X that have pixel values $I(p)$ at most t. In what is called the *greyscale filtration* of the image, we range t across all pixel values 0 to 255, take $I(p)$ to be the grayscale value of pixel p, and keep track of how and at which pixel values connected components and loops appear. The value at which a given feature appears is called its *birth* value and the value at which it merges (or closes up) is called its *death* value. The collection of birth and death values of all the features within the 0- or 1-dimensional persistent homology of I is called its (0- or 1-dimensional) *persistence diagram*.

Recent work has established the utility of incorporating multiple types of topological and geometric information derived from persistence diagrams of various filtrations of images to improve predictive capabilities of machine learning models.[19] Two such filtrations are the height[19,20] and greyscale[21-24] filtrations. In the case of the height filtration an image is first binarized at some fixed threshold and the evolution of shape is examined as the black pixels are unveiled along on a given direction in the image—see **Figure 1**. The height filtration is motivated by theoretical results that establish a two- or three-dimensional shape is completely characterized by the persistence diagrams of its height filtrations.[25] Expanding upon this result, it was established that for a 2-dimensional binary image only 4 directions are needed.[21]

Model Predictors

To derive our topological features, both height and greyscale filtrations were used. Images were smoothed before applying the greyscale filtration and before binarizing and applying the height filtration. This ensured the spatial extent of a given feature within an image was accounted for (see Section 5.2 of Reininghaus et al[26]), rather than just the "depth" of a local minima. In sum, 180 persistence diagrams in dimensions 0 and 1 were derived from the height filtration in 5 directions from grayscale images smoothed according to a Gaussian filter at 3 different bandwidths, and then binarized at 6 different thresholds. Each of these 180 diagrams was converted into a univariate summary by taking its persistent entropy.[27,28]

A total of 910 radiomics features were extracted for each image, wherein the features were constructed according to that of a previous study.[29] Finally, the trimmed mean `**tmean**` and trimmed variance `**tvar**` of pixel intensities, the aspect ratio `**ratio**` of images and two additional persistence statistics `**alps_grey**` and `**alps_bd**` were derived from each image. The ALPS statistic[24] `**alps_grey**` was calculated for the greyscale filtration after smoothing the image with $\sigma = 2$. To reduce the risk of a saturated model, the 910 radiomics features and 180 topological features were projected onto their first two principal components, denoted as `**rad_pca1**`, `**rad_pca2**`, `**tda_pca1**`, and `**tda_pca2**` respectively.

Model Development
All possible models ($2^9 - 1 = 511$) were evaluated, with a default value of regularization $C = 1$, (the definition of which can be seen in the scikit-learn[30] documentation). The five best performing models were selected according to different predictive variables (**Table 1**). The first 3 models consisted of all features of a given type: the complete "All Predictors" model, the "Radiomics only" model (all 5 such features), the "TDA only" model (all 4 such features) and the two best models (criteria seen below) consisting of subsets of the predictors in the "All Predictors" and "Radiomics only" models respectively. Logistic regression with an l^2 penalty was used because of its stable objective function [31] with respect to input changes and the decreased likelihood of observing overfitting. The (unlabeled) dataset was defined as $X = (X_1, \dots, X_{18})$. Models were validated using leave-one-out cross-validation to estimate the area under the curve (AUC) of the ROC curve[32]. The predicted probabilities for the ROC curve $\pi_1, \dots, \pi_{18}$ were defined with π_i as the predicted probability of a follicular carcinoma for patient i, based on a model trained on all observations besides i.

Results

Evaluation of Model Performance
The best performing model comprising both TDA and Radiomics features included `**tmean**`, `**alps_bd**`, `**tda_pca1**`, and `**tda_pca2**` and achieved an AUC of 0.88, a sensitivity of 0.91, and a specificity of 0.57 (for a threshold of 0.5). This was denoted as "Best Multimodal" model. The best model without any TDA features included the predictor `**ratio**` and yielded an AUC of 0.68, a sensitivity of 0.91, and a specificity of 0.43. This was denoted as "Best Radiomics" model. Corresponding AUCs for the "All Predictors", "TDA only", and "Radiomics only" models were 0.74 (sensitivity 0.91, specificity 0.43), 0.82 (sensitivity 0.91, specificity 0.71), and 0.49 (sensitivity 0.73, specificity 0.43) respectively.

For most models except "Best Multimodal" the optimal amount of regularization was around 1 (**Figure 2**), demonstrating the "Best Multimodal" model's strength by choosing λ such that its maximum AUC was not achieved. Therefore, the regularization parameter was fixed at its original value $\lambda = 1$. Given that λ tended towards 0 from its moderate values—meaning overfitting or large parameter values were not penalized—the Best Multimodal, All Predictors, and Best Radiomics models did not decline in terms of AUC. The other models exhibited a sort of unimodal structure to these model selection curves, suggesting that the multimodal models were capturing *legitimate* associations within the dataset.

Model Validation
Subsampling bootstrap was used to evaluate the uncertainty of each AUC, sensitivity, and specificity for the five models.[33] Sensitivity and specificity were calculated with prediction threshold 0.5. To estimate the sampling distribution of each metric, 1000 random samples of size $b = 10$ were uniformly chosen at random from the set of all subsets where 4 patients had adenomas and 6 had carcinomas, and then each metric was evaluated on said sample. This approach, as opposed to a naïve bootstrap method, ensured that no duplicate observations appeared in our example, so that the bootstrap samples did not have the opportunity to train on examples it would be tested on. Unlike a conventional asymptotic confidence interval, the values from which these intervals were calculated can be considered as from the correct image distribution. The values of Leave-one-out AUC, sensitivity, and specificity for each of the 5 main models along with their 95% confidence intervals can be seen in **Figures 3–5** and **Table 2**. Confidence intervals for the AUCs using DeLong's test[34] are listed in **Table 3**.

Discussion

We developed a multimodal ML classifier model incorporating both TDA and radiomics to generate predictions of carcinoma versus adenoma from ultrasound images of thyroid follicular lesions. Our "Best Multimodal" model achieved an AUC of 0.88, outperforming the models including only radiomics and

only TDA. To our knowledge, this is the first study to apply TDA to thyroid ultrasound images to distinguish follicular carcinoma from adenoma, and demonstrates the potential of topological features to not only augment the predictive power of ML classifier models, but capture legitimate distinctions between follicular carcinoma and adenoma.

We hypothesize that the main reason that radiomics performed so poorly is because it is data driven. Best practices suggest 10-15 samples per radiomics feature.[35] This puts radiomics at a severe disadvantage. Additionally, many features within the "ATA Nodule Sonography Pattern Risk of Malignancy"[3] can be described topologically. For example in the greyscale filtration, a cyst would show low (0-dimensional) persistence entropy as it consists of a single, uniform dark region; a spongiform nodule would contain many significant 0-dimensional connected components in its persistence diagram. Furthermore, microcalcifications would present as many significant holes in the 1-dimensional greyscale filtration persistence diagram. As previously mentioned, irregular margins (via notions of fractal dimension[36]) and generic shape features (such as an acute angle interface, using the height filtration)[21,25] can be captured topologically. Nonetheless, many traditional radiomics features can often be described simply so their inclusion can assist with interpretability and serve to capture relevant pixel statistics that the TDA features only capture indirectly (such as statistics relating to contrast).

There is also the question of which predictors we chose to include in our analysis. The radiomics features we chose for this study were the mean and variance of pixel intensities, the aspect ratio of the image, and the dimensionality-reduced radiomics features from a previous study[29]. The first two are fundamental statistical summaries and the aspect ratio shows up in the ATA sonography patterns associated with malignancy[3]. The topological features we chose were known throughout the literature to either correspond to relevant nodule patterns (`**alps_grey**` and `**alps_bd**`) or capture shape effectively (`**tda_pca1**` and `**tda_pca2**`). Submodels were then chosen according to the inclusion of the various radiomics and TDA features or by model performance, evaluated by AUC. Once again, summaries of the submodels predictive capacities can be seen in **Figures 2 and 3**.

Notable in Figure 3 is the fact that the interval for the AUC of the "Radiomics only" model contains 0.5, meaning that there is no evidence to conclude the model performs better than random guessing. Additionally, there is minimal overlap between AUCs of the Best Multimodal and Best Radiomics models, suggesting that with additional data a multimodal model including TDA features might be *statistically significantly* better than a model containing only Radiomics features to classify follicular carcinoma on ultrasound. Two other points are also of interest. It seems that the "All Predictors" model performs sub-optimally with respect to AUC in Figure 3 and specificity in Figure 5. The likely reason for this is a degree of overfitting—there are 9 predictors and only 18 observations on which to train and test our models. Another phenomenon that seems to be present is that the overfitting is being induced by the radiomics features, of which there are a total of 5 in said model. Finally, the 95% interval for specificity for the "TDA only" model does not contain 0.5, suggesting it is the only model that is learning the "benign" features effectively. Of course, this conclusion is advisory due to low volume data and no correction for multiple testing.

Our ability to evaluate thyroid nodules using ultrasound is currently limited by the lack of a standardized approach to manual interpretation as well as significant interobserver variability. As a means to improve the diagnostic power of thyroid ultrasound, multiple artificial intelligence (AI) techniques have been applied to attempt automated stratification of thyroid nodules. The majority employ ML models that are tasked with classifying nodules on ultrasound as benign versus malignant. Our "Best Multimodal" ML model resulted in AUC of 0.88, sensitivity of 0.91 and specificity of 0.57. While these results are at first glance inferior to prior ML classifier models that included all thyroid nodules, this could be attributed to

the particular difficulty of interpretation and diagnosis of follicular thyroid lesions and the lower volume dataset .[37]

Few studies using ML have focused specifically on follicular subtypes in thyroid nodules. Shin et al investigated use of radiomics to develop a ML classifier model to differentiate between follicular adenoma and carcinoma using ultrasound images, demonstrating high specificity and comparable diagnostic performance in comparison to two experienced radiologists.[17] Our group's prior proof-of-concept study developed a multimodal ML model that incorporated both clinical perioperative data with radiomic features extracted from ultrasound images to predict follicular adenoma from carcinoma, and showed modest improved predictive power in comparison to image alone or clinical variable alone models.[29] Given that follicular thyroid neoplasms pose a particular diagnostic dilemma within thyroid pathology, the ability of innovative techniques such as ML and TDA to enhance pre-operative classification of these lesions and subsequently prevent unnecessary surgeries will continue to be an area of focus.

In the field of medical imaging analysis, TDA has been previously applied to computed tomography (CT) and magnetic resonance imaging (MRI), in an attempt to capture and extract information from these images to allow for automated diagnosis or stratification.[38] Specifically, previous studies have looked to utilize TDA summaries of medical images to stratify COPD, detect COVID-19 infection, and detect osteoarthritis in earlier stages, showing potential for better performance over manual as well as gold standard machine learning models.[39-41] In a similar setting to this one, TDA, Radiomics, and multimodal models were contrasted in the context of lung tumor histology prediction[18]. However, there is still significant work to be done in validating, improving access to and encouraging the widespread use of TDA.

Our study contained several limitations. First, our study included a small number of patients, and it will be important for future studies to use large multicenter datasets to develop models that are generalizable to the larger population. Second, our patient population often are referred with ultrasound images from different institutions and machine types, and thus, there was significant variability in the quality of the images. Additionally, we did not include other follicular pattern thyroid lesions, such as non-invasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) or follicular variant papillary thyroid cancer. Lastly, we did not perform comparison of our model's performance to any currently available radiologic RSS (i.e. TI-RADS), and cannot specifically draw conclusions about our models' performance in comparison to these RSS in our patient population.

By focusing on novel topological patterns, TDA can capture features from data that are potentially missed by less interpretable methods of imaging analysis. This novel methodology may augment current clinical decision-support tools for thyroid US interpretation. And importantly, advances in this field may assist with reducing the significant variability in interobserver interpretation of thyroid ultrasounds and may improve the standardization of thyroid nodule management.

References

1. Guth S, Theune U, Aberle J, et al. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. European Journal of Clinical Investigation 2009;39(8):699-706

2. Tessler FN, Middleton WD, Grant EG, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. Journal of the American College of Radiology 2017;14(5):587-595
3. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 2015;26(1):1-133
4. Chng CL, Tan HC, Too CW, et al. Diagnostic performance of ATA, BTA and TIRADS sonographic patterns in the prediction of malignancy in histologically proven thyroid nodules. Singapore Medical Journal 2018;59(11):578-583
5. Kim J-S, Kim BG, Stybayeva G, et al. Diagnostic Performance of Various Ultrasound Risk Stratification Systems for Benign and Malignant Thyroid Nodules: A Meta-Analysis. Cancers 2023;15(424):1-13
6. Huang B, Ebner SA, Jasnit S M, et al. A Multidisciplinary Head-to-Head Comparison of American College of Radiology Thyroid Imaging and Reporting Data System and American Thyroid Association Ultrasound Risk Stratification Systems. The Oncologist 2020;25(398-403
7. Kuo T-C, Wu M-H, Chen K-Y, et al. Ultrasonographic features for differentiating follicular thyroid carcinoma and follicular adenoma. Asian Journal of Surgery 2020;43(1):339-46
8. Castellana M, Piccardo A, Virili C, et al. Can ultrasound systems for risk stratification of thyroid nodules identify follicular carcinoma? Cancer Cytopathology 2020;128(4):250-9
9. Lin J-D, Chao T-C. Follicular thyroid carcinoma: From diagnosis to treatment. Endocrine Journal 2006;53(4):441-8
10. Louis JDS, Leight GS, Tyler DS. Follicular neoplasms: The role for observation, fine needle aspiration biopsy, thyroid suppression, and surgery. Seminars in Surgical Oncology 1999;16(1):5-11
11. Sillery JC, Reading CC, Charboneau JW, et al. Thyroid Follicular Carcinoma: Sonographic Features of 50 Cases. Neuroradiology/Head and Neck Imaging 2010;194(44-54
12. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278(2):563-77
13. Edelsbrunner H, Harer JL. Computational topology: an introduction. American Mathematical Society: 2010.
14. Bukkuri A, Andor N, Darcy IK. Applications of Topological Data Analysis in Oncology. Frontiers in Artificial Intelligence 2021;13(4):eCollection 2021
15. Munch E. A User's Guide to Topological Data Analysis. Journal of Learning Analytics 2017;4(2):47–61, doi:10.18608/jla.2017.42.6
16. Liang J, Huang X, Hu H, et al. Predicting Malignancy in Thyroid Nodules: Radiomics Score Versus 2017 American College of Radiology Thyroid Imaging, Reporting and Data System. Thyroid 2018;28(8):1024-1033
17. Shin I, Kim YJ, Han K, et al. Application of machine learning to ultrasound images to differentiate follicular neoplasms of the thyroid gland. Ultrasonography 2020;39(257-265
18. Vandaele R, Mukherjee P, Selby HM, et al. Topological data analysis of thoracic radiographic images shows improved radiomics-based lung tumor histology prediction. Patterns 2023;4(1):
19. Garin A, Tauzin G. A Topological "Reading" Lesson: Classification of MNIST using TDA. 2019.

20. Oner D, Garin A, Koziński M, et al. Persistent Homology with Improved Locality Information for more Effective Delineation. IEEE Transactions on Pattern Analysis and Machine Intelligence 2023;1-8, doi:10.1109/TPAMI.2023.3246921
21. Betthauser LM. Topological Reconstruction of Grayscale Images. University of Florida: United States -- Florida; 2018.
22. Chung YM, Hu CS, Lawson A, et al. Topological approaches to skin disease image analysis. 2018.
23. Rieck B, Yates T, Bock C, et al. Uncovering the topology of time-varying fMRI data using cubical persistence. Advances in neural information processing systems 2020;33(6900-6912
24. Thomas AM, Crozier PA, Xu Y, et al. Feature Detection and Hypothesis Testing for Extremely Noisy Nanoparticle Images using Topological Data Analysis. Technometrics 2023;1-14, doi:10.1080/00401706.2023.2203744
25. Turner K, Mukherjee S, Boyer DM. Persistent homology transform for modeling shapes and surfaces. Information and Inference: A Journal of the IMA 2014;3(4):310-344
26. Reininghaus J, Kotava N, Günther D, et al. A scale space based persistence measure for critical points in 2d scalar fields. IEEE Transactions on Visualization and Computer Graphics 2011;17(12):2045-2052
27. Rucco M, Castiglione F, Merelli E, et al. Characterisation of the idiotypic immune network through persistent entropy. Springer: 2016.
28. Atienza N, Gonzalez-Díaz R, Soriano-Trigueros M. On the stability of persistent entropy and new summary functions for topological data analysis. Pattern Recognition 2020;107(107509
29. Lin AC, Liu Z, Lee J, et al. Generating a Multimodal Artificial Intelligence Model to Differentiate Benign and Malignant Follicular Neoplasms of the Thyroid: A Proof-of-Concept Study. Surgery 2023;In press(
30. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011;12(2825--2830
31. Shalev-Shwartz S, Ben-David S. Understanding machine learning: From theory to algorithms. Cambridge University Press: 2014.
32. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006;27(8):861-874, doi:https://doi.org/10.1016/j.patrec.2005.10.010
33. Politis DN, Romano JP, Wolf M. Subsampling. Springer Science & Business Media: 1999.
34. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;837-845
35. Shur JD, Doran SJ, Kumar S, et al. Radiomics in oncology: a practical guide. Radiographics 2021;41(6):1717-1732
36. Jaquette J, Schweinhart B. Fractal dimension estimation with persistent homology: a comparative study. Communications in Nonlinear Science and Numerical Simulation 2020;84(105163
37. Ludwig M, Ludwig B, Mikuła A, et al. Cancers. The Use of Artificial Intelligence in the Diagnosis and Classification of Thyroid Nodules: An Update 2023;15(708):1-24
38. Skaf Y, Laubenbacher R. Topological data analysis in biomedicine: A review. Journal of Biomedical Informatics 2022;130(104082):

39. Belchi F, Pirashvili M, Conway J, et al. Lung Topology Characteristics in patients with Chronic Obstructive Pulmonary Disease. Scientific Reports 2018;8(5341):
40. Iqbala S, Ahmed HF, Qaiser T, et al. Classification of COVID-19 via Homology of CT-SCAN. arXiv:210210593 2021;
41. Pedoia V, Haefeli J, Morioka K, et al. MRI and biomechanics multidimensional data analysis reveals R2-R1ρ as an early predictor of cartilage lesion progression in knee osteoarthritis. Journal of Magnetic Resonance Imaging 2017;47(1):78-90

Figures

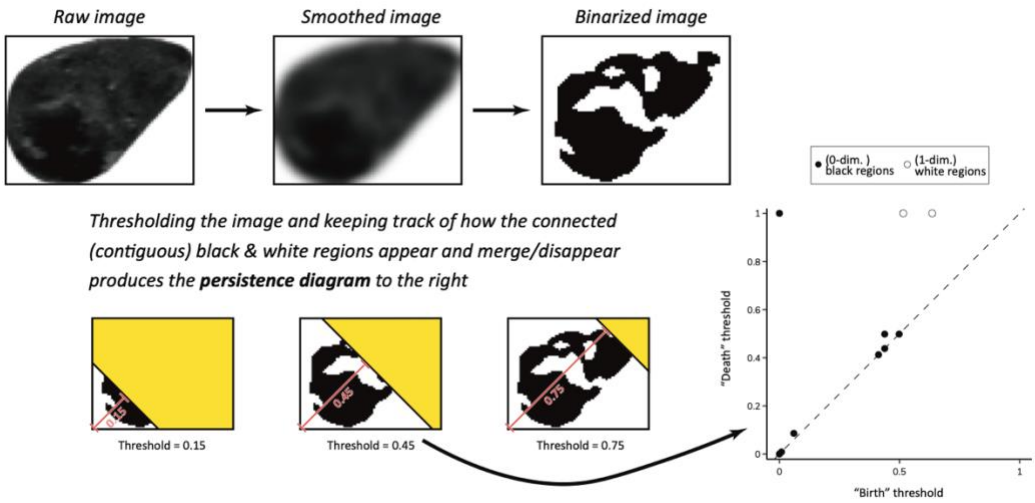

Figure 1: An illustration of the height filtration and its derived persistence diagram(s) for an ultrasound image of a follicular adenoma.

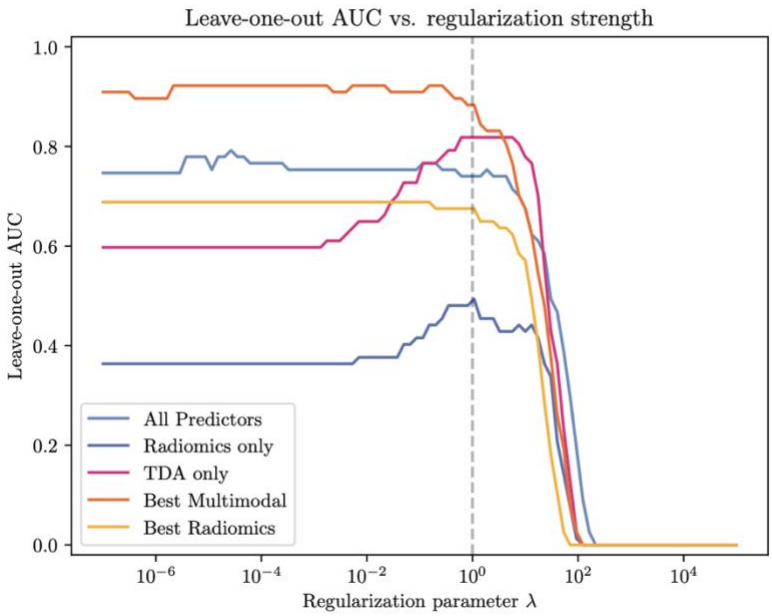

Figure 2: Plot of leave-one-out AUC vs. regularization strength for each of the five main models considered in this paper. Vertical line indicates where regularization parameter equals 1.

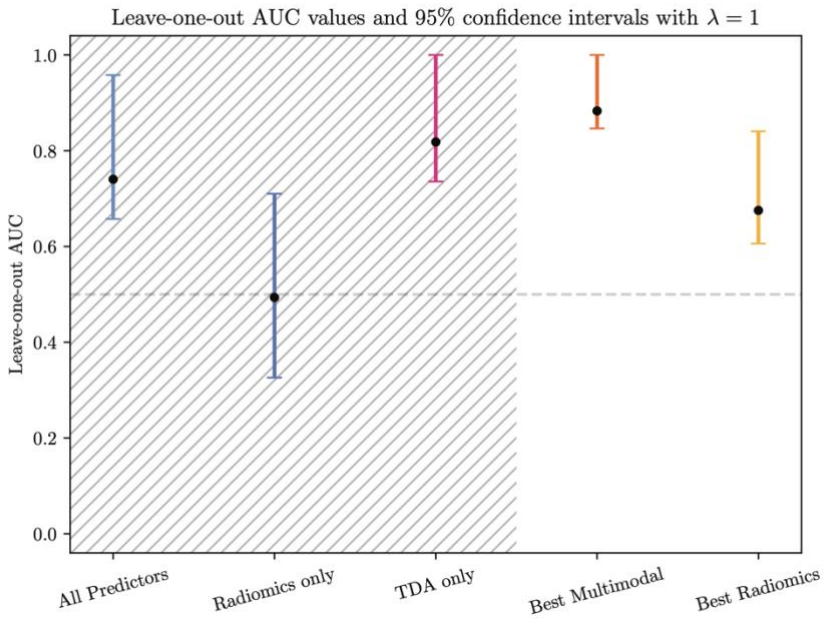

Figure 3: 95% subsampling bootstrap (marginal) confidence intervals for the AUC of each of the five main models considered in this paper. Shaded region corresponds to "full" models; unshaded region corresponds to the best models after model selection based on leave-one-out AUC.

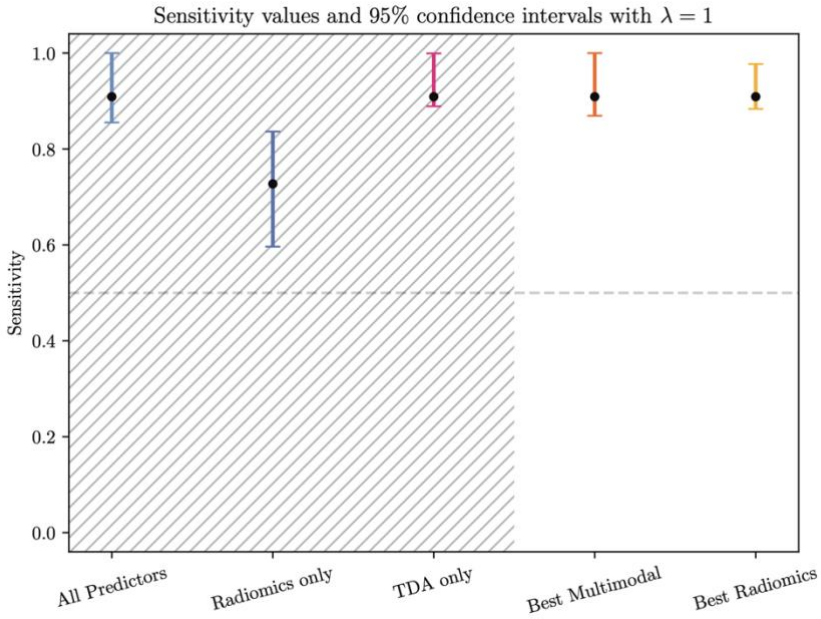

Figure 4: 95% subsampling bootstrap (marginal) confidence intervals for the sensitivity of each of the five main models considered in this paper. Shaded region corresponds to "full" models; unshaded region corresponds to the best models after model selection based on leave-one-out AUC.

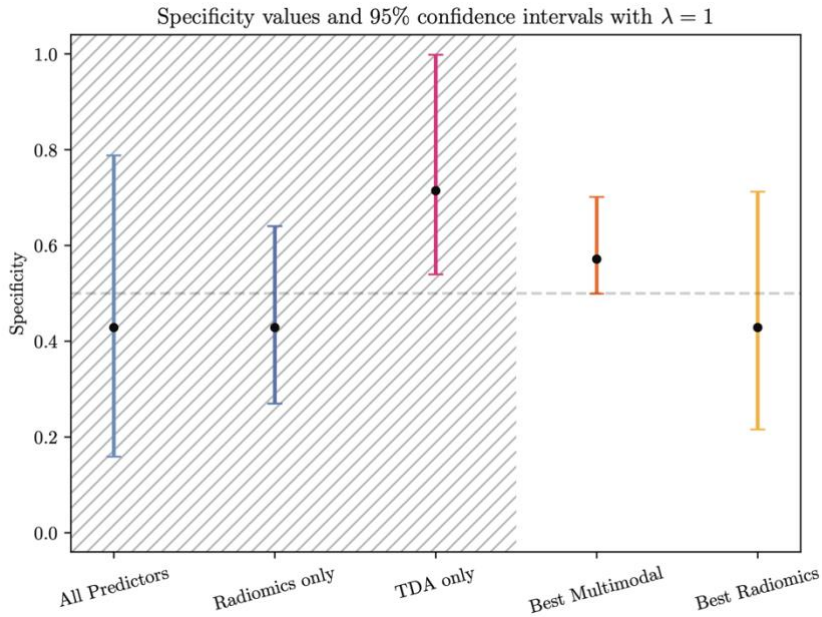

Figure 5: 95% subsampling bootstrap (marginal) confidence intervals for the specificity of each of the five main models considered in this paper. Shaded region corresponds to "full" models; unshaded region corresponds to the best models after model selection based on leave-one-out AUC.

Tables

Table 1. *List of predictors included in each of the five main models of the paper.*

	All Predictors	Radiomics only	TDA only	Best Multimodal	Best Radiomics
tmean	X	X		X	
tvar	X	X			
ratio	X	X			X
alps_grey	X		X		
alps_bd	X		X	X	
rad_pca1	X	X			
rad_pca2	X	X			
tda_pca1	X		X	X	
tda_pca2	X		X	X	

Table 2. *95% bootstrap (marginal) confidence intervals for the AUC, sensitivity, and specificity of each of the five main models considered in this paper, as well as seen in Figures 2–4.*

Even though this table only contains marginal confidence intervals, elementary probability calculations yield 90% (rectangular) confidence regions. For example, with 90% confidence, the values of (sensitivity, specificity) for the TDA only model lie in the rectangle $[0.89,1] \times [0.54,1]$. *In other words, with 90% confidence, the TDA only model has specificity at least 0.89 and specificity 0.54 at threshold equal to 0.5.*

	All Predictors	**Radiomics only**	**TDA only**	**Best Multimodal**	**Best Radiomics**
AUC	0.74 [0.66, 0.96]	0.49 [0.33, 0.71]	0.82 [0.74, 1]	0.88 [0.85, 1]	0.68 [0.61, 0.84]
Sensitivity	0.91 [0.86, 1]	0.73 [0.60, 0.84]	0.91 [0.89, 1]	0.91 [0.87, 1]	0.91 [0.88, 0.98]
Specificity	0.43 [0.16, 0.79]	0.43 [0.27, 0.64]	0.71 [0.54, 1]	0.57 [0.50, 0.70]	0.43 [0.22, 0.71]

Table 3. *95% asymptotic (marginal) confidence intervals for AUC using DeLong's test.*

	All Predictors	**Radiomics only**	**TDA only**	**Best Multimodal**	**Best Radiomics**
AUC	0.74 [0.49, 1]	0.49 [0.16, 0.83]	0.82 [0.59, 1]	0.88 [0.72, 1]	0.68 [0.38, 0.97]