

# **Breast Cancer Detection with Topological Machine Learning**

Ankur Yadav University of Texas at Dallas Richardson, TX, USA ankur.yadav@utdallas.edu Nisha Yadav University of Texas at Dallas Richardson, TX, USA nxn180012@utdallas.edu Baris Coskunuzer University of Texas at Dallas Richardson, TX, USA coskunuz@utdallas.edu

## ABSTRACT

Screening for breast cancer using mammograms and ultrasound images is an essential but time-consuming and expensive process that requires a trained clinician's interpretation. To address this issue, machine learning (ML) methods have been developed in recent years as clinical decision-support tools. However, most of these algorithms face challenges related to computational feasibility, reliability, and interpretability. We present a new approach for feature extraction in mammograms and ultrasound images using topological data analysis (TDA) methods. The proposed method uses persistent homology to capture distinct topological patterns in healthy and unhealthy patient images, which are then transformed into powerful feature vectors. These vectors are combined with standard ML techniques to create the Topo-BRCA model, which provides competitive results with state-of-the-art deep learning (DL) models in several benchmark datasets. Unlike most DL models, Topo-BRCA does not require data augmentation or preprocessing and is effective for both small and large datasets. Additionally, the topological feature vectors can easily be integrated into future DL models to enhance their performance further.

#### **CCS CONCEPTS**

• Computing methodologies → Machine learning; Machine learning; • Applied computing → Health informatics; Life and medical sciences; Health care information systems.

#### **KEYWORDS**

Breast Cancer Diagnosis, Topological Data Analysis, Mammogram, Ultrasound, Machine Learning, Cubical Persistence

#### **ACM Reference Format:**

Ankur Yadav, Nisha Yadav, and Baris Coskunuzer. 2023. Breast Cancer Detection with Topological Machine Learning. In 2023 10th International Conference on Biomedical and Bioinformatics Engineering (ICBBE 2023), November 09–12, 2023, Kyoto, Japan. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3637732.3637744

#### **1** INTRODUCTION

Breast cancer is a global health problem with a significant impact on women's health, as it is the most prevalent cancer and leads to more disability-adjusted life years (DALYs) lost than any other type of cancer [39]. In 2020 alone, 2.3 million women were diagnosed with



This work is licensed under a Creative Commons Attribution International 4.0 License.

*ICBBE 2023, November 09–12, 2023, Kyoto, Japan* © 2023 Copyright held by the owner/author(s). ACM ISBN 979-8-4007-0834-3/23/11 https://doi.org/10.1145/3637732.3637744 breast cancer, resulting in 685,000 deaths worldwide. Additionally, 7.8 million women who were previously diagnosed with breast cancer were alive at the end of 2020. Early diagnosis is crucial for the successful treatment of breast cancer, and mammography is a commonly used screening method to detect it [12]. However, due to limitations such as a small dataset, long pre-processing times, and the lack of interpretability in decision-making, the existing machine learning (ML) algorithms for breast cancer detection have not been implemented at the clinical stage.

To address this challenge, we propose using topological data analysis (TDA) techniques. TDA has been effective in feature extraction and can reveal hidden patterns in data. By identifying relevant topological features and combining them with ML models, we can create more reliable and transparent models for breast cancer detection. Our proposed approach aims to improve early diagnosis, which is crucial for successful treatment. In 2020 alone, 2.3 million women were diagnosed with breast cancer, leading to 685,000 deaths worldwide. Our work seeks to contribute to reducing these statistics and improving women's health globally.

Our proposed method uses TDA to identify distinct topological patterns in breast ultrasound and mammogram images that distinguish healthy and cancerous images (Figure 4 and 5). The main tool used in TDA, persistent homology, converts these patterns into highly effective feature vectors that, when paired with suitable ML models, produce outstanding results in benchmark datasets (See Table 2). Our model is computationally efficient and robust, does not require any data augmentation or pre-processing, and can easily be combined with deep learning methods to boost their performance.

#### **1.1 Our contributions:**

We study a novel approach to breast cancer diagnosis by applying the latest topological data analysis methods.

- By studying the evolution of topological patterns in mammogram and ultrasound images, we observe that normal and abnormal images produce distinct topological patterns (Figure 4 and 5).
- With our unique topological feature extraction method, our computationally feasible model gives competitive results in detecting breast cancer for mammograms, it falls behind for ultrasound images.
- With our powerful topological descriptors, our proposed model is highly explainable and interpretable (Section 4.2).
- Our topological feature vectors provide a key ingredient for any future ML and DL models in the domain to boost their performance and improve robustness.

### 2 RELATED WORK

### 2.1 ML in Breast Cancer Detection

In recent years there has been significant enhancement in ML tools and they have been widely employed in breast cancer image classification. In [43] they work on Deep Learning assisted Efficient Adaboost Algorithm (DLA-EABA) for breast cancer detection with advanced computational techniques. This study starts with examining the CNN-based transfer learning to characterize breast masses for different diagnostic, predictive tasks or prognostic or in several imaging modalities, such as Magnetic Resonance Imaging (MRI), Ultrasound (US), digital breast tomosynthesis and mammography. The deep learning framework contains several convolutional layers, LSTM, Max-pooling layers.

On the other hand, in [2], the authors proposed the Full resolution Convolutional Network (FrCN) with a CAD framework for X-ray mammograms. To detect the mass as malignant or benign and classify it. The publicly accessible and annotated INbreast database was used for the calculation of the suggested integrated CAD framework in terms of accuracy of classification, identification, and segmentation.

Similarly, in [9] they study the consequence of transfer learning and by experimentation establish the fine-tuning tactic to implement when working out a CNN model. They fine-tuned some of the recent, most powerful CNNs and achieved better results compared to other state-of-the-art methods which classified the same public datasets. After pre-processing and normalizing all the obtained Regions of Interest (ROIs) from the full mammograms, they combined all the datasets to create one large dataset of images and applied it to fine-tune the CNNs. In [27] new CAD system is proposed for classifying benign and malignant mass tumors in breast mammography images. The deep convolutional neural network (DCNN) architecture named AlexNet is used and is fine-tuned to classify two classes. The last fully connected (fc) layer is connected to the support vector machine (SVM) classifier to obtain better accuracy on publicly available datasets (1) the digital database for screening mammography (DDSM); and (2) the Curated Breast Imaging Subset of DDSM (CBIS-DDSM).

There are several other works on breast cancer diagnosis with deep learning methods. For a thorough review and comparison of these approaches, see excellent surveys [34], [17], [31]

### 2.2 TDA in Image Analysis

Persistent homology (PH) has been quite effective for pattern recognition in image and shape analysis in the past two decades. In medical image analysis, PH produced power results in cell development [22], tumor detection [10], neuronal morphology [19], brain artery trees [5], fMRI data [30], and genomic data [6].

See the excellent survey [36] for a thorough review of TDA methods in biomedicine. For a collection of TDA applications in several domains, see TDA Applications Library [16]. For a nice introduction to applications of TDA in biology, see the book on the subject [26].

#### **3 BACKGROUND**

#### 3.1 Persistent Homology

Persistent homology is a mathematical concept and computational tool used in topological data analysis (TDA) to analyze and understand the shape and structure of complex data sets. In general, it can be applied to various forms data, including point clouds, graphs, and images. PH allows us to systematically assess the evolution of various hidden patterns in the data as we vary a scale parameter [8]. While PH is a very effective data mining method for many data types (e.g., point clouds, networks), here we only summarize PH setup in image data case, in particular *cubical persistence*. For details of other forms of data, see [13].



Figure 1: In the left, we have original grayscale mammogram image X. Next three images  $X_0, X_{50}$  and  $X_{100}$  are binary images in the filtration where in  $X_k$ , all pixels with color value  $\leq k$ are activated (black) where all others are not activated (white).  $\beta_0(X_k)$  and  $\beta_1(X_k)$  represent the number of components and loops in  $X_k$  respectively.

In practice, PH machinery is a 3-step process. For a given image X (say  $r \times s$  resolution), the first step is to create a nested sequence of binary images (aka cubical complexes). To create such sequence, one can use grayscale (or other color channels) values  $\gamma_{ii}$  of each pixel  $\Delta_{ii} \subset X$ . In particular, for a sequence of grayscale values  $(t_1 < t_2 < \cdots < t_N)$ , one obtains a nested sequence of binary images  $X_1 \subset X_2 \subset \cdots \subset X_N$  such that  $X_n = \{\Delta_{ij} \subset X \mid \gamma_{ij} \leq t_n\}$ (See Figure 1). In other words, we start with empty  $r \times s$  image, and start activating (coloring black) pixels when their grayscale value reach the given threshold. This is called *sublevel filtration* for Xwith respect to given function (grayscale in this case). Then in the second step, PH captures the evolution of topological features in this sequence, and records as persistence diagram (PD). In particular, if a topological feature  $\sigma$  first appears in  $X_m$  and disappears in  $X_n$ with  $1 \le m < n \le N$ , we call  $b_{\sigma} = t_m$  birth time and  $d_{\sigma} = t_n$  the *death time* of the topological feature  $\sigma$ . Then, PD is the collection of all such 2-tuples  $PD_k(X) = \{(b_{\sigma}, d_{\sigma})\}$  where k represent the dimension of the topological features. The difference  $d_{\sigma} - b_{\sigma}$  is called *lifespan* of the topological feature.

PDs being collection of 2-tuples are not very practical to be used with ML tools. Instead, a common way is to convert PD information into a vector or a function, called *vectorization* [13], which is the final step of PH process. A common function for this purpose is the *Betti function*, which basically keeps track of the number of "alive" topological features at the given threshold. In particular, the Betti function is a step function with  $\beta_0(t_n)$  the count of connected components in the binary image  $X_n$ , and  $\beta_1(t_n)$  the number of Breast Cancer Detection with Topological Machine Learning



Figure 2: Sublevel filtration. For an image X of  $5 \times 5$  size with the given pixel values, the sublevel filtration is the sequence of binary images  $X_1 \subset \cdots \subset X_5$ .

holes (loops) in  $X_n$ . In ML applications, Betti functions are usually taken as a vector  $\vec{\beta}_k$  of size N with entries  $\beta(t_n)$  for  $1 \le n \le N$ . i.e.,  $\vec{\beta}_k(X) = [\beta_k(t_1) \ \beta_k(t_2) \ \beta_k(t_3) \ \dots \ \beta_k(t_N)]$ . e.g., for the image Xin Figure 2, we have  $\vec{\beta}_0(X) = [3\ 2\ 1\ 1\ ]$  and  $\vec{\beta}_1(X) = [0\ 2\ 1\ 2\ 0]$ , e.g.,  $\beta_0(1) = 3$  is the count of components in  $X_1$  and  $\beta_1(4) =$ 2 is the count of holes (loops) in  $X_4$ . There are several ways to convert PDs into a vector, e.g. Persistence landscapes, Persistence Images, Silhouettes [13], but to keep the model interpretable, we specifically chose to use Betti functions in this work as others are not easy to interpret.

### **4 BREAST CANCER DIAGNOSIS WITH TDA**

#### 4.1 Topo-BRCA for Breast Cancer Detection

In the flowchart (Figure 3), we summarized our machine learning model. For a given breast cancer ultrasound or mammogram image X, we first get its grayscale image. Then, by constructing a sublevel filtration with a grayscale function, we obtain the persistence diagrams  $PD_0(X)$  and  $PD_1(X)$ . As explained in Section 3.1, the filtration is nothing but a sequence of black-white figures where the dark points represent the pixels with a grayscale value less than the given threshold (Figure 2). Then, the persistence diagram  $PD_0(X)$  is the summary of 0-dimensional topological features (connected components in the figures in the sequence), and  $PD_1(X)$  s the summary of 1-dimensional topological features (loops in the figures in the sequence). Then, we induce functions (topological summaries) out of these persistence diagrams to use ML tools more effectively.

In this paper, we mainly used one type of topological summaries to convert the induced persistence diagrams from breast cancer ultrasound or mammogram images to *topological fingerprints*, i.e., induced feature vectors/functions as a unique identifier of an input image. These vectorization methods with different parameters gave us two topological fingerprints for each input image, i.e., Betti-0 and Betti-1 functions. There are other vectorization methods to be used in this setting (3.1), but to keep the discussion focused; we only used these two topological fingerprints. For details of Betti functions.

We employed the Random Forest and XGBoost algorithms in the machine-learning portion of our model. All the 200 features can currently be used as input for these methods. However, initially, with PCA, we tried to perform dimensionality reduction, but our models did not perform well. We used feature significance values generated by the models to make feature selection. One can perform this process manually by looking into the Betti confidence bands. However, to fully automate the process, feature selection from the model is used to choose the most important characteristics and remove collinearity between features, which increases the performance of the ML model.

The next step is to apply ML tools to these topological fingerprints. For each dataset, we applied different ML methods like Random Forest and XGBoost to these topological fingerprints of the chest X-ray images in the dataset for the classification problem (Benign/Malignant classes). The feature selection helps a lot in fine-tuning the models as it reduces the computation time. In Tables 2, 3, and 4, we give the performance of several variations of our methods obtained by different ML models on our benchmark datasets.

## 4.2 Explainability and Interpretability of Topological Fingerprints

As mentioned in the introduction, one of the main advantages of our model is explainability and interpretability. In Figure 4 and 5, we illustrate the topological patterns created by each class in Breast Mammogram and Ultrasound images. In these figures, we give median curves and 40% confidence bands of each class for the corresponding dataset. To obtain our median curves and confidence bands we used a common method called *nonparametric confidence band for the median* [15].

In Figure 4 and 5, we observe that our topological feature vectors (Betti-0 and Betti-1 vectors) successfully distinguish different classes in Mammogram and Ultrasound images. As mentioned before while other vectorization methods for PH are hard to interpret, we use Betti functions as vectorization methods to keep the model interpretable. Recall that in grayscale, the value 0 represents black, and 255 represents white. In our experiments, we renormalized [0, 255] interval to [0, 100], hence we get 100-dimensional Betti-0 and Betti-1 vectors. In the figures, x-axis represent the grayscale value in [0, 100] (0 is black, 100 is white). For Betti-0 curves, the y-axis represents the count of components, and for Betti-1 curves, y-axis represents the count of holes/loops. Hence, for an ultrasound image X, for grayscale value  $t \in [0, 100], \beta_0(t)$  represent the number of components in the binary image  $X_t$  (See Figure 2). For example, in Figure 5-left, we observe that median curves and confidence bands for two classes are very different around x = 10 and x = 40, e.g.  $\beta_0(40) \sim 75$  for benign and  $\beta_0(40) \sim 125$  for malignant. This interprets benign ultrasound images at x = 40 (unnormalized grayscale value 100) have much more components than the benign class. In other words, benign binary images  $X_{40}$  are more spread out (disconnected) than malignant ones. Similarly, in Figure 4-right, we have Betti-1 curves for a mammogram (CC view) images. For benign class  $\beta_1(50) \sim 350$  while for malignant class  $\beta_1(50) \sim 200$ . This means benign mammogram images have much more holes (loops) than the malignant ones at x = 50 (unnormalized grayscale value 125), i.e., benign binary images  $X_{50}$  have more holes than the malignant ones (Figure 2).

From ML perspective, these figures prove how strong our feature vectors are. For any image, we get 100-dimensional Betti-0, and Betti-1 vectors. The thinness/thickness of the confidence bands represents how well the given class forms a cluster in  $\mathbb{R}^{100}$ . In our case, one can consider that the median curves represent the center of the cluster for each class, and the feature vector of each image



Figure 3: Flowchart of our Topo-BRCA model: For any input image. we get their persistence diagrams by using these pixel values and obtain their topological features (Betti curves). Along with other features, we employ standard ML tools (RF, XGBoost), which provide highly accurate results for breast cancer diagnosis.



Figure 4: Breast Mammogram: In the figures above, we give the median curves and 40% confidence bands of our topological feature vectors (Betti functions) for each class in CBIS-DDSM datasets. *x*-axis represents grayscale values (renormalized from [0,255] to [0,100]) and *y*-axis represents count of components (Betti-0) or count of loops (Betti-1).

in that class lands somewhere nearby. The separation of median curves and confidence bands represents the distance between the clusters in the latent (feature) space.



Figure 5: Breast Ultrasound: In the figures above, we give the median curves and 40% confidence bands of our topological feature vectors (Betti functions) for each class in Breast Ultrasound datasets. *x*-axis represents grayscale values (renormalized from [0,255] to [0,100]) and *y*-axis represents count of components (Betti-0) or count of loops (Betti-1).

### **5 EXPERIMENTS**

#### 5.1 Datasets

**BUSI Dataset** was compiled in 2018 from Baheya Hospital for Early Detection and Treatment of Women's Cancer in Cairo, Egypt is publicly available [3]. As one of the few publicly available datasets for breast ultrasounds, the dataset serves as benchmark dataset for different ML models in the domain (Table 4).

**CBIS-DDSM** is a revised and standardized edition of DDSM dataset [21]. It involves a subset of the DDSM data selected and curated by a trained mammographer, revised mass segmentation and bounding boxes, and pathologic diagnosis for training data, configured similarly to modern computer vision data sets. It serves as a benchmark dataset for ML models for mammograms (Table 2). The statistical details of the datasets are given in Table 1.

Table 1: Summary Statistics of Benchmark datasets for Breast Cancer images

Dataset	Total	Benign	Malignant	Normal
BUSI [3]	780	210	437	133
CBIS-DDSM [21]	1696	912	784	0

#### 5.2 Experimental Setup

We give the details of our datasets in 1. Note that the majority of datasets do not have a predefined train-test split. This is why many models used their *train:test* split, as it can be seen from our accuracy tables (Tables 2,3 and 4). We used 5 and 10-fold cross-validation in this study. To ensure a fair comparison, the basic details of each method are provided in the accuracy tables. The study did not use data augmentation as the topological features extracted are invariant under rotations or flips and perform well on small and unbalanced datasets.

We have used different configurations of classes for classification. For the ultrasound image dataset, we have done 3-class, as well as binary classification. We have used only binary classification (benign vs. malignant) for the Mammogram dataset as they provided only two classes. Breast Cancer Detection with Topological Machine Learning

We first obtained each image's topological fingerprints (TFs), generated 200 features, and used feature selection to select the 50 most important features for the model. This is achieved using importance weights from the model trained with the default configuration. We selected 50 features with the highest importance values. We used *select\_from\_model* method from the python-sklearn package [25] to obtain these features. Then, we obtained our accuracy results by applying ML models (RF, KNN, XGBoost, QDA) and fine-tuning these models to these TFs. In the tables below, we only reported our best results. In Table 5, we give the performances of all variations of our models with different combinations of topological fingerprints and ML models.

In this expriment, for 500 tiles, it takes < 8 minutes for 200 feature extraction on the system with Intel Core i5-9600K and 16GB Memory (No GPU). We used Giotto-TDA [37] to obtain persistence diagrams and Betti functions. The code is given at https://anonymous.4open.science/r/TOPO-BRCA-F415.

#### 5.3 Results

We give the comparison of the performance of our Topo-BRCA model with state-of-the-art methods in Tables 2, 3,4. In Table 2, we observe that our topological approach gives highly accurate results for mammogram images by outperforming most of the SOTA DL models. Considering that we use no data augmentation or preprocessing, this high performance in such a small dataset show the robustness of our model. It also indicates that topological feature vectors can be very helpful to obtain a reliable clinical-decision support method for mammogram images.

Table 2: Accuracy results for tumor diagnosis from mammogram images on CBIS-DDSM dataset. Note that train-test splits are different.

Mammogram Results							
Method	Ben:Mal	Train:Test	Sen	Spec	Acc	AUC	
DensetNet [7]	912:784	80:20	-	-	87.5	93.0	
Unet [38]	912:784	5 fold CV	-	-	77.6	86.2	
VGG [35]	912:784	85:15	86.1	80.1	-	91.0	
InceptionV3 [1]	912:784	5 fold CV	-	98.0	88.8	-	
MVFF [24]	912:784	78:22	81.8	72.0	77.6	84.0	
DCNN [28]	912:784	5 fold CV	86.2	87.7	87.2	94.0	
FusionMR8 [42]	912:784	70:30	89.9	97.9	94.3	97.0	
2D V-net [29]	912:784	80:20	95.0	-	97.0	98.0	
Topo-BRCA	912:784	5 fold CV	94.4	94.4	94.3	97.3	

Table 3: Accuracy results for tumor diagnosis from ultrasound images on BUSI dataset. Note that train-test splits are different.

Ultrasound Binary Results						
Method	Ben:Mal	Train:Test	Sen	Spec	Acc	AUC
SAC [18]	109:54	10 fold CV	88.2	78.7	-	-
BVA Net [40]	210:437	90:10	75.8	88.3	84.3	88.9
MCWT [32]	210:437	10 fold CV	94.3	97.7	96.6	-
Topo-BRCA	210:437	10 fold CV	79.3	77.2	79.3	84.8

On the other hand, our experiments show that topological approach does not work very well in ultrasound images (Table 3 and

Table 4: Accuracy results for tumor diagnosis from ultrasound images on BUSI dataset for three classes. Note that train-test splits are different.

Ultrasound 3-class Results							
Method	Ben:Mal:Nor	Train:Test	Sen	Spec	Acc	AUC	
SAC [18]	210:437:133	10-Fold CV	82.5	62.8	-	-	
ViT [14]	210:437:133	85:15	-	-	83.0	92.0	
BGWO [20]	210:437:133	10-Fold CV	-	-	84.9	97.0	
ResNet-18 [11]	210:437:133	10-Fold CV	-	-	88.9	-	
BUViTNet [4]	210:437:133	90:10	-	-	91.3	93.7	
HAHCC [33]	130:130:130	5-Fold CV	-	92.7	91.7	92.3	
CNN [23]	210:437:133	80:20	83.7	90.3	88.4	90.5	
RMTL-Net [41]	210:437:133	5-Fold CV	93.3	86.1	91.0	96.7	
Topo-BRCA	210:437:133	5 Fold CV	68.6	79.2	68.4	77.5	

4). One of the possible reasons for this performance is the mixed resolution images in BUSI dataset [3]. The values of our Betti functions depends on the resolution, and mixed size images can produce highly different Betti functions for the same class.

*5.3.1 Ablation Study.* In our ablation study, we study the relative performance of our topological feature vectors (Betti functions) in different dimensions (Table 5). Our results indicate that combining all dimensions improves the results in general, and

Table 5: Comparison of the performances of different topological feature vectors for BUSI dataset. Accuracy results are given in %.

Ablation Study						
Input Features	3 Class	B vs M	N vs M	B vs N		
100 Betti-0	63.3	74.6	78.1	79.4		
100 Betti-1	65.6	74.0	80.7	81.4		
200 Betti Features	68.1	77.2	79.8	82.4		

## 6 CONCLUSION

Breast cancer is the most prevalent type of cancer in women next to lung cancer, and early detection significantly increases the survival rate. Screening for breast cancer using mammograms and ultrasound images is an essential but time-consuming and expensive process that requires a trained clinician's interpretation. In this work, we studied a novel approach to this problem by applying the latest topological data analysis tools. Our computationally efficient model has yielded highly competitive results when compared to the latest deep learning models in mammogram screening. It's noteworthy, however, that similar methods don't yield the same success when applied to ultrasound images. Considering that mammograms remain the primary mode of breast cancer screening, we anticipate that our topological features will emerge as invaluable resources in the years of research and study ahead. Furthermore, when fused with state-of-the-art deep learning models, our distinctive topological feature vectors have the potential to play a pivotal role in the development of exceptionally precise and robust topological deep learning models, thus addressing this crucial necessity more effectively. Moving forward, our forthcoming studies will focus on this specific direction.

ICBBE 2023, November 09-12, 2023, Kyoto, Japan

#### REFERENCES

- Richa Agarwal, Oliver Diaz, Xavier Lladó, Moi Hoon Yap, and Robert Martí. 2019. Automatic mass detection in mammograms using deep convolutional neural networks. *Journal of Medical Imaging* 6, 3 (2019), 031409–031409.
- [2] Mugahed A Al-Antari, Mohammed A Al-Masni, Mun-Taek Choi, Seung-Moo Han, and Tae-Seong Kim. 2018. A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *International journal of medical informatics* 117 (2018), 44–54.
- [3] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. 2020. Dataset of breast ultrasound images. Data in brief 28 (2020), 104863.
- [4] Gelan Ayana and Se-woon Choe. 2022. BUViTNet: Breast Ultrasound Detection via Vision Transformers. *Diagnostics* 12, 11 (2022), 2654.
- [5] Paul Bendich, James S Marron, Ezra Miller, Alex Pieloch, and Sean Skwerer. 2016. Persistent homology analysis of brain artery trees. *The annals of applied statistics* 10, 1 (2016), 198.
- [6] Pablo G Cámara, Arnold J Levine, and Raul Rabadan. 2016. Inference of ancestral recombination graphs through topological data analysis. *PLoS computational biology* 12, 8 (2016), e1005071.
- [7] Eduardo Castro, Jose Costa Pereira, and Jaime S Cardoso. 2023. Symmetry-based regularization in deep breast cancer screening. *Medical Image Analysis* 83 (2023), 102690.
- [8] F. Chazal and B. Michel. 2021. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in Artificial Intelligence* 4 (2021).
- [9] Hiba Chougrad, Hamid Zouaki, and Omar Alheyane. 2018. Deep Convolutional Neural Networks for breast cancer screening. *Computer Methods and Programs* in Biomedicine 157 (2018), 19–30. https://doi.org/10.1016/j.cmpb.2018.01.011
- [10] Lorin Crawford, Anthea Monod, Andrew X Chen, Sayan Mukherjee, and Raúl Rabadán. 2020. Predicting clinical outcomes in glioblastoma: an application of topological and functional data analysis. *J. Amer. Statist. Assoc.* 115, 531 (2020), 1139–1150.
- [11] Arijit Das and Srinibas Rana. 2021. Exploring residual networks for breast cancer detection from ultrasound images. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 1–6.
- [12] Carol E DeSantis, Stacey A Fedewa, Ann Goding Sauer, Joan L Kramer, Robert A Smith, and Ahmedin Jemal. 2016. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. CA: a cancer journal for clinicians 66, 1 (2016), 31–42.
- [13] Tamal Krishna Dey and Yusu Wang. 2022. Computational topology for data analysis. Cambridge University Press.
- [14] Behnaz Gheflati and Hassan Rivaz. 2022. Vision transformers for classification of breast ultrasound images. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 480–483.
- [15] Jean Dickinson Gibbons and Subhabrata Chakraborti. 2014. Nonparametric statistical inference. CRC press.
- [16] Barbara Giunti. 2022. TDA Applications Library. https://www.zotero.org/groups/ 2425412/tda-applications/library.
- [17] Ghada Hamed, Mohammed Abd El-Rahman Marey, Safaa El-Sayed Amin, and Mohamed Fahmy Tolba. 2020. Deep learning in breast cancer detection and classification. In Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020). Springer, 322–333.
- [18] Kuan Huang, Yingtao Zhang, H. D. Cheng, and Ping Xing. 2021. Shape-Adaptive Convolutional Operator for Breast Ultrasound Image Segmentation. In 2021 IEEE International Conference on Multimedia and Expo (ICME). 1–6. https://doi.org/10. 1109/ICME51207.2021.9428287
- [19] Lida Kanari, Paweł Dłotko, Martina Scolamiero, Ran Levi, Julian Shillcock, Kathryn Hess, and Henry Markram. 2018. A topological representation of branching neuronal morphologies. *Neuroinformatics* 16, 1 (2018), 3–13.
- [20] Priyanka Khanna, Mridu Sahu, and Bikesh Kumar Singh. 2021. Improving the classification performance of breast ultrasound image using deep learning and optimization algorithm. In 2021 IEEE International Conference on Technology, Research, and Innovation for Betterment of Society (TRIBES). IEEE, 1–6.
- [21] Rebecca Sawyer Lee, Francisco Gimenez, Assaf Hoogi, and Daniel Rubin. 2016. Curated breast imaging subset of DDSM. *The cancer imaging archive* 8 (2016), 2016.
- [22] Melissa R McGuirl, Alexandria Volkening, and Björn Sandstede. 2020. Topological data analysis of zebrafish patterns. *Proceedings of the National Academy of Sciences* 117, 10 (2020), 5113–5124.
- [23] Woo Kyung Moon, Yan-Wei Lee, Hao-Hsiang Ke, Su Hyun Lee, Chiun-Sheng Huang, and Ruey-Feng Chang. 2020. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Computer Methods and Programs in Biomedicine* 190 (2020), 105361. https://doi.org/10.1016/j.cmpb.2020.105361
- [24] Hasan Nasir Khan, Ahmad Raza Shahid, Basit Raza, Amir Hanif Dar, and Hani Alquhayz. 2019. Multi-View Feature Fusion Based Four Views Model for Mammogram Classification Using Convolutional Neural Network. *IEEE Access* 7 (2019), 165724–165733. https://doi.org/10.1109/ACCESS.2019.2953318

- [25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [26] Raúl Rabadán and Andrew J Blumberg. 2019. Topological data analysis for genomics and evolution: topology in biology. Cambridge University Press.
- [27] D. A. Ragab, M. Sharkas, S. Marshall, P. M. C. 6354665. Ren j. breast cancer detection using deep convolutional neural networks, and support vector machines. peerj. 2019 jan 28;7:e6201. doi: 10.7717/peerj.6201. pmid: 30713814; pmcid:PMC6354665. [n. d.]. ([n. d.]).
- [28] Dina A Ragab, Maha Sharkas, Stephen Marshall, and Jinchang Ren. 2019. Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ* 7 (2019), e6201.
- [29] Khalil ur Rehman, Jianqiang Li, Yan Pei, Anaa Yasin, Saqib Ali, and Yousaf Saeed. 2021. Architectural distortion-based digital mammograms classification using depth wise convolutional neural network. *Biology* 11, 1 (2021), 15.
- [30] Bastian Rieck, Tristan Yates, Christian Bock, Karsten Borgwardt, Guy Wolf, Nicholas Turk-Browne, and Smita Krishnaswamy. 2020. Uncovering the topology of time-varying fMRI data using cubical persistence. Advances in neural information processing systems 33 (2020), 6900-6912.
- [31] Roslidar Roslidar, Aulia Rahman, Rusdha Muharar, Muhammad Rizky Syahputra, Fitri Arnia, Maimun Syukri, Biswajeet Pradhan, and Khairul Munadi. 2020. A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection. *IEEE Access* 8 (2020), 116176–116194.
- [32] Tariq Sadad, Ayyaz Hussain, Asim Munir, Muhammad Habib, Sajid Ali Khan, Shariq Hussain, Shunkun Yang, and Mohammed Alawairdhi. 2020. Identification of Breast Malignancy by Marker-Controlled Watershed Transformation and Hybrid Feature Set for Healthcare. *Applied Sciences* 10, 6 (2020). https://doi.org/ 10.3390/app10061900
- [33] Adyasha Sahu, Pradeep Kumar Das, and Sukadev Meher. 2023. High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets. *Biomedical Signal Processing and Control* 80 (2023), 104292.
- [34] Li Shen, Laurie R Margolies, Joseph H Rothstein, Eugene Fluder, Russell McBride, and Weiva Sieh. 2019. Deep learning to improve breast cancer detection on screening mammography. *Scientific reports* 9, 1 (2019), 12495.
- [35] Li Shen, Laurie R Margolies, Joseph H Rothstein, Eugene Fluder, Russell McBride, and Weiva Sieh. 2019. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. *Scientific Reports* 9, 1 (Aug. 2019), 12495.
  [36] Yara Skaf and Reinhard Laubenbacher. 2022. Topological Data Analysis in
- [36] Yara Skaf and Reinhard Laubenbacher. 2022. Topological Data Analysis in Biomedicine: A Review. Journal of Biomedical Informatics (2022), 104082.
- [37] Guillaume Tauzin, Umberto Lupo, Lewis Tunstall, Julian Burella Pérez, Matteo Caorsi, Anibal Medina-Mardones, Alberto Dassatti, and Kathryn Hess. 2020. giotto-tda: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration. arXiv:2004.02551 [cs.LG]
- [38] Lazaros Tsochatzidis, Panagiota Koutla, Lena Costaridou, and Ioannis Pratikakis. 2021. Integrating segmentation information into CNN for breast cancer diagnosis of mammographic masses. *Computer Methods and Programs in Biomedicine* 200 (2021), 105913.
- [39] WHO. 2021. Breast Cancer. https://www.who.int/news-room/fact-sheets/detail/ breast-cancer/. [Online; accessed 26-march-2021].
- [40] Jie Xing, Chao Chen, Qinyang Lu, Xun Cai, Aijun Yu, Yi Xu, Xiaoling Xia, Yue Sun, Jing Xiao, and Lingyun Huang. 2020. Using BI-RADS stratifications as auxiliary information for breast masses classification in ultrasound images. *IEEE Journal of Biomedical and Health Informatics* 25, 6 (2020), 2058–2070.
- [41] Meng Xu, Kuan Huang, and Xiaojun Qi. 2023. A Regional-Attentive Multi-Task Learning Framework for Breast Ultrasound Image Segmentation and Classification. *IEEE Access* (2023).
- [42] Qian Zhang, Yamei Li, Guohua Zhao, Panpan Man, Yusong Lin, and Meiyun Wang. 2020. A novel algorithm for breast mass classification in digital mammography based on feature fusion. *Journal of Healthcare Engineering* 2020 (2020), 1–11.
- [43] Jing Zheng, Denan Lin, Zhongjun Gao, Shuang Wang, Mingjie He, and Jipeng Fan. 2020. Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis. *IEEE Access* 8 (2020), 96946–96954.