Topology-Informed Flood Detection from Satellite Images

Max Zhao

1 Personal Story

This project emerged from two questions that at first seemed unrelated. The first was a machine learning question: how can a computer learn from images that are not clean, centered, and simple? The second was a mathematical question: what does it mean for a shape to stay the "same" when it is stretched, bent, blurred, or partly hidden? My research began when I realized that floods sit at the intersection of both questions.

During my junior year, I took an artificial intelligence class and encountered neural networks near the end of the course. I had seen machine learning described in broad terms before, but this was the first time I began to understand why the subject was so powerful. A neural network is not given a list of explicit rules. Instead, it learns patterns from many examples. In image problems, modern networks can learn small visual features such as edges, textures, and repeated local patterns. That idea immediately appealed to me because it felt like a bridge between mathematics, computer science, and the physical world.

At the same time, I have always been drawn to problems that begin with a very simple question but refuse to stay simple. In solving puzzles, I enjoyed the moment when a familiar object became strange. A Sudoku puzzle is just a grid until one asks why it is hard to solve. A satellite image is just a picture until one asks what information is hidden inside each pixel. A river is just a river until one asks how a machine could recognize that it has overflowed.

In the summer of 2025, I joined the Science and Engineering Apprenticeship Program and worked at the U.S. Naval Research Laboratory under Dr. Dandan Li. Dr. Li works on deep learning analysis of radar images. She did not hand me a finished project with a fixed method. Instead, she introduced me to the field of remote sensing and gave me room to explore. Because she knew I was interested in mathematics, she gave me some papers on topological data analysis, or TDA. I began reading about persistent homology, one of the main tools of TDA, and I was fascinated by its central promise: it can describe features of data that remain stable under small perturbations.

Topology is often introduced through the joke that a coffee mug and a doughnut are the same because each has one hole. That joke contains a serious idea. Topology studies properties that survive continuous deformation. A circle can become an oval and remain topologically the same. A pond can become slightly wider, darker, or blurrier in an image, but it may still have the same broad connected structure. This made me wonder whether topology could help with remote sensing, where satellite images are rarely perfect.

I first got my hands dirty by using TDA libraries on the MNIST handwritten digit dataset. MNIST is a classic machine learning benchmark made of small grayscale images of handwritten digits. It is tidy compared with real satellite data, but it was a useful place to learn how persistent homology turns an image into a collection of mathematical features. After that, I began looking for satellite imagery to test. I was surprised to find that persistent homology had not, to the best of my literature search, been applied directly to synthetic aperture radar backscatter images. That surprise was motivating, but it also made the project more difficult. There was no standard recipe for the exact problem I wanted to solve.

Real satellite images quickly humbled me. They are not like textbook pictures. Optical images can be blocked by clouds. Radar images can see through clouds, but they contain speckle noise and other artifacts. Different satellites collect different bands at different resolutions. Even the same location can look very different on two different days for reasons unrelated to flooding. A mathematical stability theorem is helpful on paper, but it does not automatically solve these engineering problems.

Just around this time, severe flooding struck central Texas in July 2025. The news made the project feel much less abstract. Floods, from a math perspective, are changes in connected bodies of water. A river widens, tributaries merge, fields become submerged, and disconnected patches may appear. These are exactly the kinds of patterns where the overall shape of an image may matter more than a single local texture. I began to see flood detection as a natural setting for topology.

Dr. Li helped me find a review paper on flood detection with synthetic aperture radar, which emphasized the need for reproducible benchmarks and carefully chosen datasets [1]. With limited access to large GPU systems, I needed a project that was ambitious but still possible to complete in one summer. I chose the open-source SEN12-FLOOD dataset, which contains time series of satellite images from flood events and non-flood periods [2]. Instead of trying to draw the exact boundary of every flooded region, I focused on the earlier and lighter-weight task of detecting whether each frame in a time sequence showed flooding.

My work was full-time for nine weeks, from June 16 to August 15, 2025. Most of the planning and analysis happened at the Naval Research Laboratory. Because I was a minor and did not have full access to some lab computing systems, much of the model training was done on my own computer at home. That limitation was frustrating at first. I could not run every experiment I wanted, and I had to be careful about model size and training time. But it also shaped the project in a useful way. It pushed me toward compact topological features and models that could train quickly.

The hardest part of the project was not a single theorem or a single line of code. It was the repeated loop of trying an idea, discovering that the data were messier than expected, and then redesigning the method. I spent weeks preprocessing the images, computing topological features, testing classifiers, and finding that straightforward approaches did not converge. Eventually, I developed an adaptive Gaussian embedding for persistence diagrams. That embedding became the bridge between topology and neural networks in the final model.

This project made science and mathematics feel more alive to me because neither field stayed in its own box. The topology was not decoration; it guided what features I should look for. The machine learning was not magic; it forced me to ask whether the features were actually useful. The satellite data were not just arrays; they represented real landscapes and real disasters. I had to move back and forth between abstraction and application, and that movement was the most exciting part.

My advice to other high school students interested in projects combining science and mathematics is to choose a problem where the mathematics has a job to do. It is tempting to start with an impressive tool and then search for a place to use it. My project began to work only when the tool and the phenomenon matched: topology studies stable shape, and flooding is a change in the shape and connectivity of water. I would also recommend keeping careful records of failures. Many of my failed attempts were not wasted; they told me what the data would not tolerate.

2 Research Story

Floods are the most common natural disaster [3, 4, 5]. Rapid flood detection matters because emergency responders need to know where to send attention, especially when roads are closed or ground reports are delayed. Satellite imagery is valuable for this task because satellites can observe large regions repeatedly. The challenge is that satellite imagery during floods is often messy.

Two kinds of satellites are especially useful. Optical satellites, such as Sentinel-2, record reflected sunlight in multiple color and infrared bands. These images are often intuitive for humans, but clouds can block the view exactly when storms and floods are happening. Synthetic aperture radar, or SAR, satellites such as Sentinel-1 send microwave signals toward Earth and measure the returned signal. SAR can image the ground day or night and through clouds, but SAR images are noisy and visually unfamiliar. The best flood detector should therefore be able to use both optical and radar information.

Another important point is that flooding is a time-series problem. A single satellite image may not be enough. A dark patch in a radar image could be open water, a wet surface, a shadow-like artifact, or something else. But if a river is narrow in earlier frames and then expands over later frames, the change is strong evidence for flooding. For this reason, previous work on the SEN12-FLOOD dataset used a convolutional neural network to extract features from each image and then a recurrent neural network called a gated recurrent unit, or GRU, to process the sequence over time [2, 6].

Convolutional neural networks, or CNNs, are excellent at learning local image patterns. They look at small neighborhoods of pixels and gradually combine those patterns into more complex features. This is powerful for many image tasks. However, floods also have global structure. The crucial signal may be that water regions become connected, expand, or form large coherent shapes. A CNN can learn some of this indirectly, but topology gives a more direct language for describing connected components and holes.

The main idea of my project is that flood detection can improve when local convolutional features and global topological features are combined. In simple terms, the CNN describes texture and local appearance, while persistent homology describes shape and connectivity. The final model uses both.

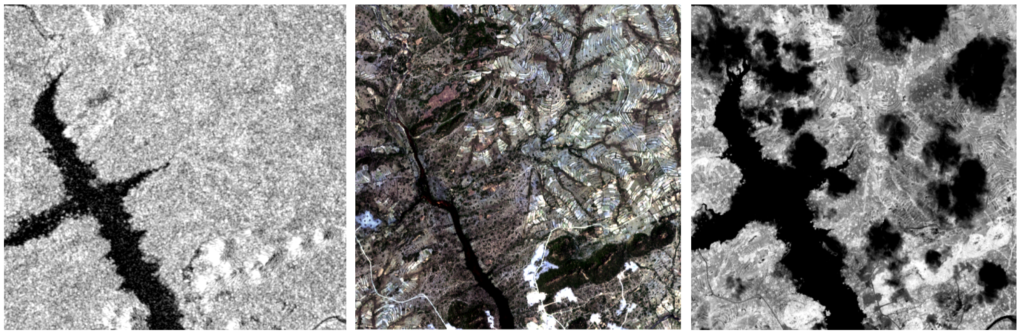

Figure 1: Examples of preprocessed satellite frames. Left: a Sentinel-1 SAR image with visible speckle. Middle: a Sentinel-2 optical image shown in RGB. Right: a grayscale map derived from optical bands using a water index. These examples show why flood detection from satellite images is not the same as ordinary clean-image classification.

To explain persistent homology, imagine slowly filling in an image according to pixel intensity. At a low threshold, only some pixels are included. As the threshold rises, more pixels appear, small islands merge, and loops can form or disappear. Persistent homology records when each connected component or loop is born and when it dies. A feature that exists for a long range of thresholds is considered more persistent and usually more meaningful than one that appears only briefly.

For a grayscale image, this process produces a persistence diagram. Each point in the diagram is a birth-death pair (b, d). The value b says when a feature appears, and d says when it disappears. The lifetime $d - b$ measures how persistent that feature is. Points far from the diagonal line $b = d$ represent long-lived features; points near the diagonal often represent noise.

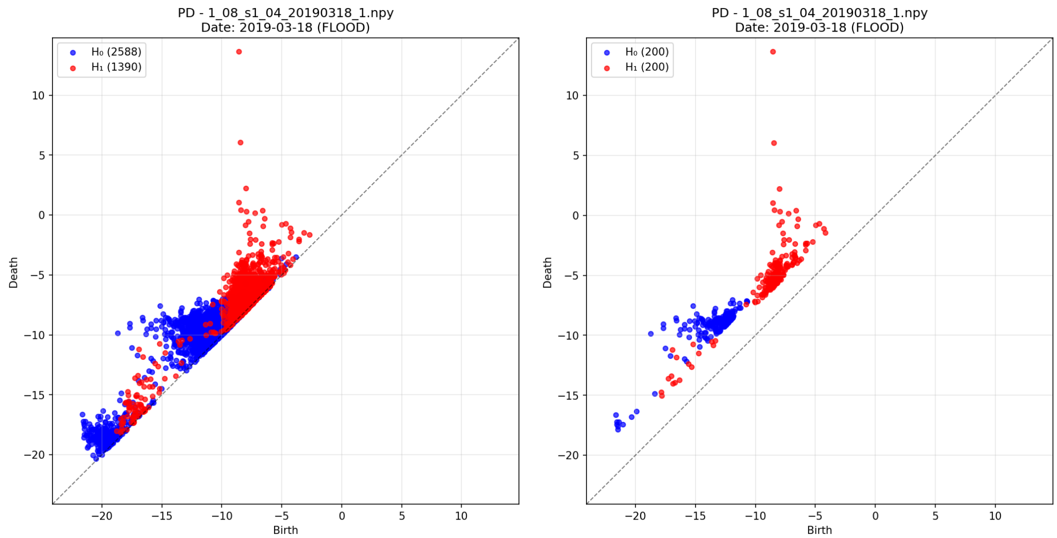

Figure 2: Persistence diagrams computed from a SAR image. Each point represents a topological feature. The right panel keeps only the most persistent features, which are more likely to describe robust structure rather than noise.

The mathematical object is elegant, but a machine learning model cannot directly consume a variable-size cloud of points. A neural network expects fixed-length numerical vectors. This was one of the main technical obstacles in my project. I tried several representations of persistence diagrams, but many did not train reliably on the satellite time series. The diagrams shifted from frame to frame because of clouds, atmospheric variation, SAR speckle, and differences between sensors.

To solve this, I developed an adaptive Gaussian grid embedding. The method first looks at all persistence diagrams from the training data and builds a grid on the birth-death plane using quantiles of observed birth and death values. This makes the grid adapt to the actual satellite data instead of imposing a fixed scale. Then, for each persistence diagram, every topological point contributes to nearby grid centers through a Gaussian weight. Longer-lived features are weighted more heavily. The resulting vector has fixed length and can be passed to a GRU.

The core formula is

$$\phi_j(D) = \sum_{(b_i,d_i)\in D} (d_i - b_i) \exp\left(-\frac{\|(b_i,d_i) - c_j\|_2^2}{2\sigma^2}\right), \tag{1}$$

where D is a persistence diagram, c_j is a grid center, and σ controls the width of the Gaussian.

The important idea is not the formula alone, but what it accomplishes: it turns a shifting set of topological points into a stable numerical fingerprint of the image.

The dataset I used, SEN12-FLOOD, contains 335 satellite image sequences. Each sequence corresponds to a geographic location in Africa, Iran, or Australia and includes Sentinel-1 SAR frames, Sentinel-2 optical frames, or both. Each frame is labeled as flood or no flood. For SAR data, I used the two radar polarization channels and converted valid pixel values to decibel scale. For optical data, I used multispectral bands for CNN features and used the Normalized Difference Water Index, or NDWI, to create a grayscale water map for topological analysis [7]. Since high NDWI indicates water while low SAR backscatter often indicates water, I negated the NDWI map so the topological filtration would be more consistent across modalities.

I implemented three models. The first, ResNet-GRU, used only convolutional features. A ResNet-50 network extracted a feature vector from each frame, and a GRU processed the frame sequence. This reproduced and improved the earlier style of model for the dataset by using transfer learning from a remote sensing dataset [8, 9].

The second model, Topo-GRU, used only topological features. Each frame was converted into a persistence diagram, embedded into a fixed-length vector, and then passed to the GRU. This model tested whether topology alone could detect flooding.

The third model, Fusion-GRU, combined both feature types. For each frame, I concatenated the ResNet feature vector with the topological feature vector. This fused descriptor was then passed to the GRU. This was the most important model because it tested the central hypothesis: local convolutional features and global topological features should complement each other.

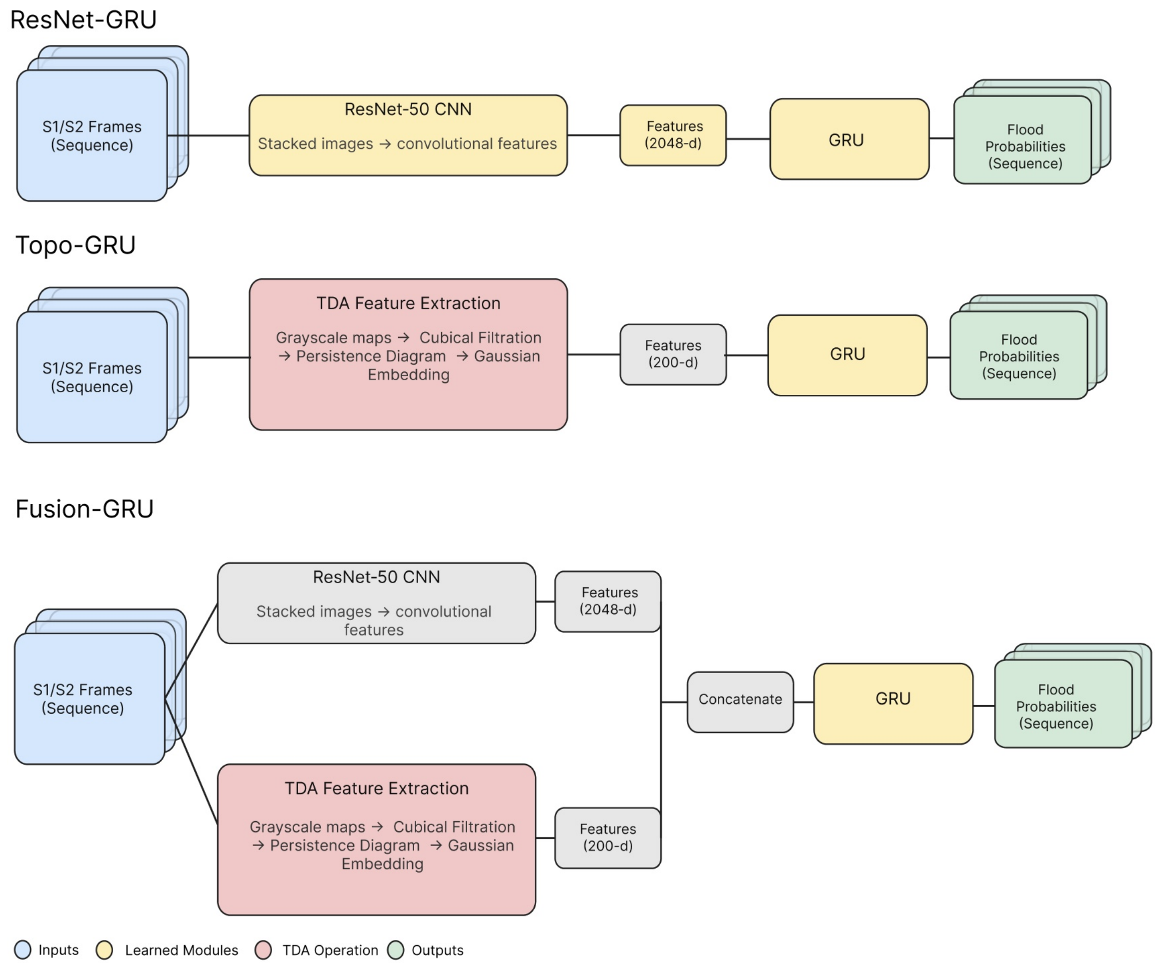

Figure 3: The three model structures used in the project. ResNet-GRU uses convolutional features, Topo-GRU uses topological features, and Fusion-GRU combines both.

The results supported the hypothesis. On dual-modality data, the previous reported accuracy for a comparable convolutional recurrent approach was 95.7% [2]. My improved ResNet-GRU reached 97.4% accuracy with transfer learning. Topo-GRU alone reached 94.1% accuracy, which was lower than the CNN-based model but its performance was impressive given that its feature vector was much smaller. The combined Fusion-GRU model achieved the best overall result, with 98.9% accuracy and strong scores on other metrics such as F_1 and recall at high precision.

Model	$\boldsymbol{F_\beta}$	$\boldsymbol{F_1}$	**Accuracy**	$\boldsymbol{R_{\mathrm{max}}}$
ResNet-GRU, dual	0.947	0.954	0.974	0.953
Topo-GRU, dual	0.916	0.904	0.941	0.903
Fusion-GRU, dual	0.980	0.982	0.989	0.986

Table 1: Selected performance results for unidirectional GRU models on dual-modality data. $R_{\max}$ is the best recall achieved while maintaining at least 90% precision.

Accuracy is useful, but flood detection is an imbalanced and safety-critical problem, so I also measured precision, recall, F_1, and an F_β score that weights recall more heavily than precision. This matters because missing a real flood can be worse than raising a false alarm. I also swept the decision threshold to find the maximum recall possible while keeping precision at or above 90%. Fusion-GRU performed well across these metrics, not just in accuracy.

The project has limitations. I trained and evaluated the models on SEN12-FLOOD because it is open source and already labeled, but future work should test on larger and more geographically diverse datasets, including locations in the United States. I also did not fully explore more advanced ways of fusing topology and CNN features, such as attention mechanisms or learnable topological layers. Because of computing constraints, I did not train a bidirectional version of Fusion-GRU or fine-tune all components jointly.

Still, the main conclusion is clear: topology can improve satellite image flood detection. Persistent homology provides a principled way to describe global structure in noisy images. CNNs offer a powerful way to describe local visual patterns. When these two views are combined, the model can detect floods more accurately than either view alone.

More broadly, this project suggests that mathematics can help machine learning see differently. The dominant approach in image analysis is to keep building larger neural networks. That approach is powerful, but it is not the only path. Sometimes a mathematical description of the phenomenon can tell the model what kind of information matters. For floods, shape and connectivity matter. For medical images, materials science, climate data, and many other noisy scientific datasets, similar topological ideas may help reveal structure that is difficult to capture from local texture alone.

To me, the most exciting part of the project is that a concept from pure mathematics became useful in a real-world problem. A flood is a physical event, a satellite image is a technological measurement, a neural network is a computational model, and persistent homology is an abstract mathematical tool. The project worked only because all four pieces clicked together. That is exactly the kind of intersection between science and mathematics that first drew me to the problem.

References

[1] Donato Amitrano, Gerardo Di Martino, Alessio Di Simone, and Pasquale Imperatore. Flood Detection with SAR: A Review of Techniques and Datasets. *Remote Sensing*, 16(4), 2024.

[2] C. Rambour, N. Audebert, E. Koeniguer, B. Le Saux, M. Crucianu, and M. Datcu. Flood detection in time series of optical and SAR images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020:1343–1346, 2020.

[3] United Nations Office for Disaster Risk Reduction (UNDRR) and Centre for Research on the Epidemiology of Disasters (CRED). The Human Cost of Disasters: An Overview of the Last 20 Years (2000–2019). Technical report, UNDRR and CRED, Geneva, Switzerland and Brussels, Belgium, 2020. `https://www.undrr.org/publication/human-cost-disasters-overview-last-20-years-2000-2019`, Accessed 2025-09-16.

[4] X. Zhang, H. Li, Y. Wu, et al. Impacts of climate warming on global floods and their implication to flood protection standards. *Journal of Hydrology*, 613:128456, 2023.

[5] IPCC. Water. In H.-O. Pörtner, D. C. Roberts, M. Tignor, et al., editors, *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, chapter 4, page 121–216. Cambridge University Press, Cambridge, UK, 2022.

[6] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, October 2014.

[7] Stuart K. McFeeters. The use of the normalized difference water index (ndwi) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7):1425–1432, 1996.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[9] Kai Norman Clasen, Leonard Hackel, Tom Burgert, Gencer Sumbul, Begüm Demir, and Volker Markl. reBEN: Refined BigEarthNet Dataset for Remote Sensing Image Analysis. *arXiv preprint: arXiv:2407.03653*, 2024.